Combinatorics, Probability and Computing (2012) 21, 916–941.
doi:10.1017/S0963548312000351

On Sums of Generating Sets in $\mathbb{Z}_2^n$

CHAIM EVEN-ZOHAR†

Einstein Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel
(e-mail: chaim.evenzohar@mail.huji.ac.il)

Received 24 August 2011; revised 4 July 2012; first published online 3 August 2012

Let A and B be two affinely generating sets of $\mathbb{Z}_2^n$. As usual, we denote their Minkowski sum by $A+B$. How small can $A+B$ be, given the cardinalities of A and B? We give a tight answer to this question. Our bound is attained when both A and B are unions of cosets of a certain subgroup of $\mathbb{Z}_2^n$. These cosets are arranged as Hamming balls, the smaller of which has radius 1.

By similar methods, we re-prove the Freiman–Ruzsa theorem in $\mathbb{Z}_2^n$, with an optimal upper bound. Denote by $F(K)$ the maximal spanning constant $|\langle A\rangle|/|A|$ over all subsets $A \subseteq \mathbb{Z}_2^n$ with doubling constant $|A+A|/|A| \leqslant K$. We explicitly calculate $F(K)$, and in particular show that $4^K/4K \leqslant F(K)\cdot(1+o(1)) \leqslant 4^K/2K$. This improves the estimate $F(K) = \mathrm{poly}(K)4^K$, found recently by Green and Tao [17] and by Konyagin [23].

AMS 2010 *Mathematics subject classification*: Primary 11P70

1. Introduction

Much work has been devoted to the study of Minkowski sums of sets. Questions concerning such sums come up in geometry, and are at the core of additive combinatorics. Research in this area has blossomed in recent years, and even Tao and Vu's monograph [36] no longer covers all the most recent developments. In this paper we concentrate on the Minkowski sum of two generating sets of $\mathbb{Z}_2^n$.

We first review some of the relevant literature. Let G be an abelian group, and let A and B be two finite subsets of G. As usual, we denote

$$A+B = \{a+b \mid a \in A, b \in B\}$$

and we ask about the minimum of $|A+B|$, given the cardinalities of A and B.

In general, the answer ranges from $\max(|A|,|B|)$ to $|A|+|B|-1$, depending on the structure of G. For a torsion-free G, if A and B are arithmetic progressions with the

† This paper is based on the author's MSc thesis, under the supervision of Professor Nati Linial.

same step, then $|A+B| = |A| + |B| - 1$, which is optimal. Likewise, if $G = \mathbb{Z}_p$ is cyclic of prime order, then the answer is given by the Cauchy–Davenport theorem, $|A+B| \geqslant \min(|A| + |B| - 1, |G|)$ [2, 4]. Moreover, by a theorem of Vosper [38], if $|A| + |B| < |G|$ then equality holds only for arithmetic progressions. In the other extreme case, G has a finite subgroup of a suitable cardinality. Thus, if $H \triangleleft G$ is a subgroup of cardinality $|H| = \max(|A|, |B|)$, an optimal choice is to have A and B be subsets of H, in which case $|A+B| = \max(|A|, |B|)$. More generally, $|A+B|$ can be as small as $\max(|A|, |B|)$ if and only if $\min(|A|, |B|) \leqslant |H|$ and $|H|$ divides $\max(|A|, |B|)$ [36, p. 55]. In the general case [8, 10], the smallest possible cardinality of $|A+B|$ is $\min(\lceil |A|/|H| \rceil + \lceil |B|/|H| \rceil - 1)|H|$, where the minimum is over all finite subgroups H of G. In a sense, this result interpolates between the two extremes. In an optimal construction [1, 10] the sets A and B are contained in $\lceil |A|/|H| \rceil$ and $\lceil |B|/|H| \rceil$ cosets of H, whose arrangement is a lexicographical variant of an arithmetic progression. In particular, for G a 2-torsion group this reduces to the well-studied Hopf–Stiefel function [20, 35, 39, 1, 7, 9].

Stability is a recurring theme in modern extremal combinatorics. Once an extremal problem is solved, it is interesting to explore what happens when we consider candidate solutions that do not resemble the global optimum. The crucial feature of the above-mentioned optimal constructions is that A and B are densely packed in cosets of properly chosen subgroups of G. We therefore return to the original question, under the requirement that A and B are not allowed to be contained in a proper subgroup of G or a coset thereof. The *affine span* of A, denoted $\langle A \rangle$, is the smallest coset (of any subgroup) containing A. We say that A *affinely generates* G if $\langle A \rangle = G$. Clearly this definition coincides with the usual notion of a generating set if $0 \in A$. The refined problem is as follows. In a finitely generated abelian group G, find $\min |A+B|$ as a function of $|A|$ and $|B|$, where A and B are finite affinely generating subsets of G.

Naturally, the structural properties of G play a role in this problem as well. For the torsion-free case, $G = \mathbb{Z}^d$, this question and similar ones were discussed by Ruzsa [30], and a full answer was finally given by Gardner and Gronchi [15]. In the extremal construction, the smaller set is a simplex of $d+1$ points, on one of whose edges lies an arithmetic progression, and the other set is roughly the sum of several copies of it. As discussed there, this is analogous to the Brunn–Minkowski theorem [33].

Here we present the following lower bound for the opposite extreme of a 2-torsion group, $G = \mathbb{Z}_2^n$.

Theorem 1.1. *Suppose $A, B \subseteq G = \mathbb{Z}_2^n$ such that $\langle A \rangle = G$, $B \neq \varnothing$ and $|A| \leqslant \frac{3}{4}|G|$. If t is the largest positive integer such that*

$$|A| \leqslant \frac{t+1}{2^t} \cdot |G|$$

and $0 \leqslant k < t$ and $w \in [-1, 1]$ are such that

$$|B| = \frac{\binom{t}{0} + \binom{t}{1} + \cdots + \binom{t}{k} + w\binom{t-1}{k}}{2^t} \cdot |G|,$$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

then

$$|A+B| \geqslant \frac{\binom{t}{0}+\binom{t}{1}+\cdots+\binom{t}{k}+\binom{t}{k+1}+w\binom{t-1}{k+1}}{2^t}\cdot|G|.$$

This bound is tight when $w=0$*, and it is attained by the sets*

$$A = D_1^t \times \mathbb{Z}_2^{n-t}, \quad B = D_k^t \times \mathbb{Z}_2^{n-t}, \quad A+B = D_{k+1}^t \times \mathbb{Z}_2^{n-t},$$

where $D_k^t = \{x \in \mathbb{Z}_2^t \mid \#\{i|x_i=1\} \leqslant k\}$ *is a Hamming ball of radius* k *in* $\mathbb{Z}_2^t$.

The Freiman–Ruzsa theorem [31] is a major result in additive combinatorics. In the context of the above discussion, it addresses the special case $A=B$. It states that if A is a subset of an r-torsion abelian group with $|A+A| \leqslant K|A|$, then A is contained in a coset of cardinality at most $F(K)|A|$, with $F(K)=K^2r^{K^4}$. The special case $r=2$ has received considerable attention [5, 6, 16, 17, 19, 23, 26, 32, 37]. Among the most recent contributions is work by Green and Tao [17] with further improvement by Konyagin [23]. It shows that one can take $F(K)=2^{2K+O(\log K)}$. Here we exactly determine the lowest possible value of $F(K)$ for $r=2$.

Theorem 1.2. *For* $K \geqslant 1$*, denote by* $t \geqslant 1$ *the unique integer for which*

$$\frac{\binom{t}{2}+t+1}{t+1} \leqslant K < \frac{\binom{t+1}{2}+(t+1)+1}{(t+1)+1}.$$

For $A \subseteq \mathbb{Z}_2^n$ *such that* $|A+A|/|A| \leqslant K$*, we have* $|\langle A\rangle|/|A| \leqslant F(K)$*, where*

$$F(K) = \begin{cases} \frac{2^t}{\binom{t}{2}+t+1}\cdot K & \frac{\binom{t}{2}+t+1}{t+1} \leqslant K < \frac{t^2+t+1}{2t}, \\ \frac{2^{t+1}}{t^2+t+1}\cdot K & \frac{t^2+t+1}{2t} \leqslant K < \frac{\binom{t+1}{2}+(t+1)+1}{(t+1)+1}. \end{cases}$$

This choice of $F(K)$ *is tight, and grows as* $\Theta(2^{2K}/K)$.

Compression is an important tool from extremal set theory. Much progress in the application of compression to additive problems was made by Bollobás and Leader in [1], and it is a key ingredient in Green and Tao's proof in [17]. There is a whole range of compression operators C that transform an arbitrary set A into another set $C(A)$, with $|C(A)|=|A|$ and $|C(A)+C(A)| \leqslant |A+A|$. By a finite sequence of such compressions, it is possible to reduce to the case where A is *compressed* in some appropriate sense, and hence has certain structural properties which make $A+A$ easier to study. The difficulty is that $C(A)$ need not be affinely generating even if A is. Green and Tao handled this difficulty by restricting the types of compression operators they used. Our approach is different. We employ more types of compression operators and we proceed as long as possible without jeopardizing affine generation, *i.e.*, as long as $\langle C(A)\rangle = \langle A\rangle$.

Isoperimetric inequalities play an important role in our work. In our investigations of $A+B$, we prove a new variant of the isoperimetric inequality for the hypercube.

Overview. In Section 2 we discuss compressions and other useful tools. We explore the key notion of compression that maintains affine generation. In Section 3 Theorem 1.2

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

is proved, first in an asymptotic form, then with the exact expression. In Section 4 we establish Theorem 1.1. The proof utilizes our new isoperimetric inequality.

2. Tools

In this section we briefly survey several concepts and results that are used below. These include the lexicographic order and the Hopf–Stiefel function. Then we discuss compressions in $\mathbb{Z}_2^n$, in line with Section 2 of [17], and we introduce the study of compressions that preserve affine generation.

2.1. The lexicographic order

Throughout, we use the linear basis $\{e_1, e_2, \ldots, e_n\}$ for $\mathbb{Z}_2^n$. Elements $x \in \mathbb{Z}_2^n$ are expressed as $x = \sum_{i=1}^n x_i e_i$. The correspondence between vectors $x \in \mathbb{Z}_2^n$ and their supports $\{j \mid x_j = 1\} \subseteq \{1, \ldots, n\} = [n]$, is used to simplify certain notation and arguments.

The *lexicographic order* is a total order on $\mathbb{Z}_2^n$. For $x, y \in \mathbb{Z}_2^n$, we say that $x \prec y$ if $x_i < y_i$ for the largest coordinate i for which $x_i \neq y_i$. For example, the ordering of $\mathbb{Z}_2^3$ is

$$0 \prec e_1 \prec e_2 \prec e_1 + e_2 \prec e_3 \prec e_1 + e_3 \prec e_2 + e_3 \prec e_1 + e_2 + e_3.$$

The *height*, $\hbar(x)$, of an element x in a finite totally ordered set is x's place in that order. For a set of elements A we denote $\hbar(A) = \sum_{x \in A} \hbar(x)$.

If $T \subseteq \mathbb{Z}_2^n$, then its *initial segment* of size a, denoted $IS(a, T)$, is the set of the a smallest elements of T in the lexicographic order. We use the abbreviation $IS(a) = IS(a, \mathbb{Z}_2^n)$ for $n \in \mathbb{N}$ large enough.

2.2. The Hopf–Stiefel function

For the reader's convenience we prove the following observation of Bollobás and Leader [1].

Proposition 2.1. *For two initial segments $IS(a), IS(b) \subseteq \mathbb{Z}_2^n$, the sum $IS(a) + IS(b)$ is an initial segment as well.*

Proof. For $z \prec x + y$, we claim that $z = x' + y'$ for some $x' \preceq x$ and $y' \preceq y$. Let $i \in \mathbb{N}$ be the largest index such that $x_i = 1$ or $y_i = 1$. Say $x_i = 1$. If $z_i = 0$, then clearly $z \prec x$, so we can take $x' = z$ and $y' = 0$. If $z_i = 1$, then note that $(z - e_i) \prec (x - e_i) + y$. By induction on i, obtain $(z - e_i) = x'' + y''$ for $x'' \preceq (x - e_i)$ and $y'' \preceq y$, and choose $x' = x'' + e_i$ and $y' = y''$. □

The Hopf–Stiefel binary function $a \circ b$ can be defined on $\mathbb{N} \times \mathbb{N}$ as follows:

$$a \circ b = |IS(a) + IS(b)|.$$

Proposition 2.1 can be restated as $IS(a) + IS(b) = IS(a \circ b)$. This definition is relevant for us for the following reason. The cardinality of a sumset of two sets of given cardinalities is minimized by taking the two sets to be initial segments:

$$a \circ b = \min\Big\{|A + B| \mid A, B \in \mathbb{Z}_2^n,\ |A| = a,\ |B| = b\Big\}.$$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

Note that here the sets are not required to be affinely generating. This result can be deduced by the technique of compressions, as we discuss below: see Lemma 2.4.

In particular, taking $A = IS(a)$ and $B = IS(b_1) \cup (e_n + IS(b_2))$ for n large enough, one can verify the sub-distributive law:

$$a \circ (b_1 + b_2) \leqslant a \circ b_1 + a \circ b_2.$$

Similarly, one can deduce the recursive relations for $a, b \leqslant 2^n$:

$$a \circ (2^n + b) = 2^n + a \circ b,$$
$$(2^n + a) \circ (2^n + b) = 2^{n+1}.$$

These two formulas can be taken as an alternative definition of the Hopf–Stiefel function [27].

The function first arose in works of Hopf [20] and Stiefel [35]. They used tools from algebraic topology to prove that $a \circ b$ provides a lower bound for solutions of the Hurwitz problem, concerning real quadratic forms (see [34]). The relation to set addition in $\mathbb{Z}_2^n$ was given by Yuzvinsky [39]. As it turns out, the Hopf–Stiefel function arises in the study of several more problems in various contexts. There is also a base-p analogue of this function for $p > 2$: see [7]. For a survey, see [9].

2.3. Compressions

For $I = \{i_1, i_2, \ldots\} \subseteq [n]$, denote $H_I = \langle 0, e_{i_1}, e_{i_2}, \ldots \rangle \lhd \mathbb{Z}_2^n$. As usual, if H is a subgroup of G, we denote by G/H the collection of all H-cosets in G. The *I-compression* of a subset $A \subseteq \mathbb{Z}_2^n$ is defined by

$$C_I(A) = \bigcup_{T \in \mathbb{Z}_2^n / H_I} IS(|A \cap T|, T).$$

In words, in every H_I-coset T we replace the elements of $A \cap T$ by a same-cardinality initial segment, with respect to the lexicographic order. We say A is compressed with respect to I, or *I-compressed*, if $C_I(A) = A$. In particular, lexicographic initial segments of $\mathbb{Z}_2^n$ are exactly all $[n]$-compressed sets.

Example 2.2. $C_{\{1,2,3\}}(\{0, e_1, e_2, e_3, e_4\}) = \{0, e_1, e_2, e_1 + e_2, e_4\}$.

This notion of compression is closely related to the operation bearing the same name from extremal set theory (see, *e.g.*, [12]). A subset of $\mathbb{Z}_2^n$ naturally corresponds to a family (a.k.a. set-system) $\mathcal{F}$ of subsets of $[n]$. We freely move between these terminologies if no confusion can occur. An $\{i\}$-compression corresponds to the *push-down operator* T_i, which replaces $J \in \mathcal{F}$ by $J \setminus \{i\}$ provided that $J \setminus \{i\} \notin \mathcal{F}$. If $\mathcal{F}$ is $\{i\}$-compressed for each i, then it is closed under taking subsets and is called a *downset*. The *shift operator* S_{ij} replaces j by i wherever possible. Namely, for every J with $i, j \notin J$ it replaces $J \cup \{j\}$ by $J \cup \{i\}$, given that the former belongs to $\mathcal{F}$ and the latter does not. We say that $\mathcal{F}$ is *shift-minimal* if it is invariant to all shifts S_{ij} where $i < j$. One can check that being $\{i, j\}$-compressed for all $i, j \in [n]$ corresponds to being a shift-minimal downset.

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

Compression can simplify matters substantially, while preserving several useful features of the set-system. Here are some observations about compressions. These and others are found in [17]. The proofs are straightforward, working coset by coset.

Lemma 2.3 (Properties of compressions). *Suppose $A \subseteq \mathbb{Z}_2^n$ and $I \subseteq [n]$.*

(i) $|C_I(A)| = |A|$.
(ii) $C_I(A)$ *is* I*-compressed.*
(iii) $\hbar(C_I(A)) \leqslant \hbar(A)$ *with equality if and only if* A *is* I*-compressed.*
(iv) *An* I*-compressed set is* J*-compressed for all* $J \subseteq I$.
(v) $C_I(A) \subseteq C_I(B)$ *for all* $A \subseteq B$. □

Compressions behave well on sumsets. By Proposition 2.1, one can deduce that the sum of two I-compressed subsets is I-compressed too. The following well-known lemma deals with the compression of a sum of two general subsets. For the sake of completeness, we prove it here, following [1] and [17].

Lemma 2.4 (Sumset compression). *Suppose $A, B \subseteq \mathbb{Z}_2^n$ and $I \subseteq [n]$. Then*

$$C_I(A) + C_I(B) \subseteq C_I(A + B).$$

Consequently $|C_I(A) + C_I(B)| \leqslant |A + B|$.

Proof. We use a double induction, on $|I|$ and on $\hbar(A) + \hbar(B)$. For the induction step, suppose that for some $J \subsetneq I$ either A or B is not J-compressed. In this case

$$\begin{aligned} C_I(A) + C_I(B) &= C_I(C_J(A)) + C_I(C_J(B)) \subseteq C_I(C_J(A) + C_J(B)) \\ &\subseteq C_I(C_J(A + B)) = C_I(A + B). \end{aligned}$$

Both inclusions are by the induction hypothesis: the first one since $\hbar(C_J(A)) + \hbar(C_J(B)) < \hbar(A) + \hbar(B)$ by property (iii) of Lemma 2.3, and the second one since $|J| < |I|$ and by property (v). The equalities are by property (iv).

It only remains to verify the lemma for A and B which are both J-compressed for all $J \subsetneq I$. We start with the simpler case $n = |I|$.

What are the subsets of $G = \mathbb{Z}_2^n$ that are J-compressed for all $J \subsetneq [n]$? By property (iv), all initial segments are such. If $S \subseteq G$ is not an initial segment, then we must have $x \notin S$ and $y \in S$ for some consecutive $x \prec y$. The only consecutive pair in G that is not contained in a proper H_J-coset is $(e_1 + \cdots + e_{n-1}) \prec e_n$. One can verify, for example by S being $[2 \cdots n]$-compressed, that the only such set is $S = H_{[n-1]} \setminus \{e_1 + \cdots + e_{n-1}\} \cup \{e_n\}$. In conclusion, it is enough to check the case where A and B are initial segments or equal to S. Now there are four cases to consider.

(1) If both A and B are initial segments, then by Proposition 2.1 $A + B$ is an initial segment too:

$$\Rightarrow C_I(A) + C_I(B) = A + B = C_I(A + B).$$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

(2) If $A = B = S$, then note that $|S| \leqslant |S + S|$ and $C_I(S) = H_{[n-1]}$:

$$\Rightarrow C_I(S) + C_I(S) = C_I(S) \subseteq C_I(S + S).$$

(3) If $B = S$ and A is an initial segment with $|A| \leqslant |S|$, then $A = C_I(A) \subseteq C_I(S) = H_{[n-1]}$:

$$\Rightarrow C_I(A) + C_I(S) = C_I(S) \subseteq C_I(A + S).$$

(4) If $B = S$ and A is an initial segment with $|A| > |S|$ then $|A| + |S| > |G|$. This means $A + S = G$, as the reader may verify by a standard pigeonhole argument:

$$\Rightarrow C_I(A) + C_I(S) = G = C_I(G) = C_I(A + S).$$

The case $n > |I|$ is implied by the case $n = |I|$:

$$\begin{aligned}
C_I(A) + C_I(B) &= \bigcup_{H_c \in G/H_I} \big(\big(C_I(A) + C_I(B)\big) \cap H_c \big) \\
&= \bigcup_{H_c \in G/H_I} \bigcup_{H_a + H_b = H_c} \big(\big(C_I(A) \cap H_a\big) + \big(C_I(B) \cap H_b\big) \big) \\
&= \bigcup_{H_c \in G/H_I} \bigcup_{H_a + H_b = H_c} \big(C_I\big(A \cap H_a\big) + C_I\big(B \cap H_b\big) \big) \\
&\subseteq \bigcup_{H_c \in G/H_I} \bigcup_{H_a + H_b = H_c} C_I\big(\big(A \cap H_a\big) + \big(B \cap H_b\big) \big) \\
&\subseteq \bigcup_{H_c \in G/H_I} C_I\Big(\bigcup_{H_a + H_b = H_c} \big(\big(A \cap H_a\big) + \big(B \cap H_b\big) \big) \Big) \\
&= \bigcup_{H_c \in G/H_I} C_I\big((A + B) \cap H_c \big) \\
&= \bigcup_{H_c \in G/H_I} \big(C_I(A + B) \cap H_c \big) \\
&= C_I(A + B).
\end{aligned}$$

The first and second inequalities are simply dividing into cases, according to the involved H_I-cosets. The third one holds because compressions work coset-wise. Then there is inclusion, by the assumption on the case $I = [n]$ applied to our H_I and translated to the relevant H_I-cosets, together with inclusion of the initial segments, because the union is at least as large as each of its components. The three remaining equalities are similar to the first three. □

2.4. Compressions that preserve affine generation

As Lemma 2.4 shows, in the problems we consider here, compressing the sets under consideration can only improve our objective function. However, we are restricting ourselves to affinely generating sets and compression may destroy this property (*e.g.*, Example 2.2). Therefore, our strategy is to keep compressing as long as affine generation is maintained. To this end we introduce the following definition.

Suppose that $A \supseteq E$, where $E = \{0, e_1, e_2, \ldots, e_n\}$ is the standard affine basis of $\mathbb{Z}_2^n$. If A is I-compressed for every I such that $C_I(A) \supseteq E$, we say that A is $\langle\langle E \rangle\rangle$-*compressed*.

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

Note that by part (iii) of Lemma 2.3 every set A containing E, can be turned into an $\langle\langle E\rangle\rangle$-compressed set by a finite sequence of such compressions. It turns out that $\langle\langle E\rangle\rangle$-compressed sets are very structured.

Lemma 2.5 (Structure of $\langle\langle E\rangle\rangle$-compressed sets). *Let $A \subseteq \mathbb{Z}_2^n$ be an $\langle\langle E\rangle\rangle$-compressed set.*

(i) *A is a shift-minimal downset.*
(ii) *A contains a subgroup of maximal size $H \triangleleft \mathbb{Z}_2^n$ of the form $H = \langle 0, e_1, \ldots, e_h\rangle$.*
(iii) *A is $\{1, \ldots, h, h+i\}$-compressed for every $1 \leqslant i \leqslant m = \operatorname{codim} H$.*
(iv) *$A \subseteq H + E$, i.e., $A = H \cup A_1 \cup A_2 \cup \cdots \cup A_m$, where $A_i = A \cap (e_{h+i} + H)$.*
(v) *For $1 \leqslant i \leqslant m$, $0 < |A_i| < |H|$.*
(vi) *For $1 \leqslant i < j \leqslant m$, $|A_i| + |A_j| \leqslant |H|$.*
(vii) *If $m > 1$, then $|A| \leqslant \left(1 + \frac{m}{2}\right)|H|$.*

Proof. The proofs are fairly straightforward.

(i) It is a simple observation that both $\{i\}$-compressions and $\{i, j\}$-compressions preserve $E \subseteq A$. Hence A must already be compressed with respect to these sets, *i.e.*, a shift-minimal downset.

(ii) Let h be the maximal dimension of a subgroup contained in A. As shown below in Lemma 2.6, a subgroup of dimension h must contain an element of Hamming weight at least h. By shift-minimality $e_1 + e_2 + \cdots + e_h \in A$, and by the downset property $H = \langle 0, e_1, \ldots, e_h\rangle \subseteq A$.

(iii) Denote $I = \{1, \ldots, h, h+i\}$. The sets $H \cup \{e_{h+i}\}$ and $\{e_{h+j}\}$ for $j \neq i$ are initial segments of their H_I-cosets. These sets cover E and remain included in A through the I-compression.

(iv) By the downset property, it is sufficient to show $e_{h+i} + e_{h+j} \notin A$ for each $1 \leqslant i < j \leqslant m$. Indeed, if A contains $e_{h+i} + e_{h+j}$ then it contains $e_{h+i} + H$ by being $\{1, \ldots, h, h+j\}$-compressed. This implies $H \cup (e_{h+i} + H) \subseteq A$, contrary to the maximality of the subgroup H in A.

(v) For the lower bound note that $e_{h+i} \in E \subseteq A$. On the other hand, if $|A_i| = |H|$ then $H \cup (e_{h+i} + H) \subseteq A$, contrary, again, to the maximality of H.

(vi) Note that $e_{h+j} \in A$, while some lexicographically smaller elements in $e_{h+i} + H$ are not contained in A. Therefore A cannot be I-compressed for $I = \{1, \ldots, h, h+i, h+j\}$. Since it is $\langle\langle E\rangle\rangle$-compressed, this means $e_{h+j} \notin C_I(A)$. Equivalently, $|A \cap H_I| \leqslant 2|H|$, which leads to our claim.

(vii) If $|A_i| \leqslant \frac{1}{2}|H|$ for every i, clearly $|A| = |H| + \sum_{i=1}^{m} |A_i| \leqslant \left(1 + \frac{m}{2}\right)|H|$. Otherwise, $|A_i| > \frac{1}{2}|H|$ for some i, thus $|A_j| \leqslant |H| - |A_i| < \frac{1}{2}|H|$ for every $j \neq i$. So $|A_i| + |A_j| \leqslant |H|$ for some i and j, and the remaining A_j are no bigger than $\frac{1}{2}|H|$. □

Lemma 2.6. *Let H be an h-dimensional subgroup of $\mathbb{Z}_2^n$. Then H contains an element of Hamming weight at least h.*

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

Proof. If $h = n$, take $e_1 + e_2 + \cdots + e_n$. Otherwise, there exists a basis element e_i such that $e_i \notin H$. In this case, moving from H to $C_{\{i\}}(H)$ simply deletes e_i from the standard basis representations of H's elements, thereby not increasing their Hamming weights. Now note that $C_{\{i\}}(H)$ is an h-dimensional subgroup of $\langle 0, e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n \rangle$, and by induction on n contains an element of Hamming weight at least h. □

3. The Freiman–Ruzsa theorem in $\mathbb{Z}_2^n$

For $A \subseteq \mathbb{Z}_2^n$ we refer to $|\langle A \rangle|/|A|$ as A's *spanning constant* and to $K = |A + A|/|A|$ as its *doubling constant*. The Freiman–Ruzsa theorem gives an upper bound on the spanning constant in terms of K. We first review the theorem and some of its quantitative aspects. Then we calculate the bound explicitly, and in particular we determine its correct asymptotics, which turns out to be $\Theta(2^{2K}/K)$. We present the proof in two stages, starting with the asymptotic estimates. We find this presentation convenient, since the proof of the asymptotic bound contains our main ideas.

3.1. Brief review of the Freiman–Ruzsa theorem

Freiman's celebrated theorem [14] states that if $A \subset \mathbb{Z}$ is a finite subset with $|A + A| \leqslant K|A|$, then A is included in a generalized arithmetic progression, whose size (relative to $|A|$) and dimension are bounded. The bounds depend only on K and not on $|A|$. Ruzsa [28, 29] has made crucial contributions to this area. More recently much work was done on similar problems where $\mathbb{Z}$ is replaced by other groups. In particular Ruzsa [31] proved the analogous result for abelian torsion groups. See [37] for a nice exposition.

Theorem 3.1 (Ruzsa). *Let G be an abelian group in which every element has order at most r. If A is a finite subset of G with $|A + A| \leqslant K|A|$, then A is contained in a coset of a subgroup $H \triangleleft G$ of size $|H| \leqslant f(r,K)|A|$, where*

$$f(r,K) \leqslant K^2 r^{K^4}.$$

Better estimates on $f(r,K)$ were subsequently found. We denote by $F(r,K)$ the smallest bound for which this statement holds. Note that $F(r,K)$ is non-decreasing in K and $F(r,1) = 1$.

By considering the case where A is an affine basis of $\mathbb{Z}_r^{2(K-1)}$, we see that $F(r,K) \geqslant r^{2K - O(\log K)}$ (see Example 3.3 below). This suggests the following conjecture [31].

Conjecture 3.2 (Ruzsa). *For some $C \geqslant 2$ we have $F(r,K) \leqslant r^{CK}$.*

In an attempt to understand the role of torsion in these phenomena, much work was dedicated to the special case $r = 2$, where $G = \mathbb{Z}_2^n$. This work is also motivated by the role that $\mathbb{Z}_2^n$ plays in discrete mathematics and in particular in coding theory [3]. We introduce the following notation:

$$F(K) = F(2,K) = \sup\left\{ \frac{|\langle A \rangle|}{|A|} \;\middle|\; A \subseteq \mathbb{Z}_2^n,\ n \in \mathbb{N},\ \frac{|A + A|}{|A|} \leqslant K \right\}.$$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

As observed by Ruzsa [5], for $r = 2$ his method gives somewhat more, namely $F(K) \leqslant K2^{\lfloor K \rfloor^3 - 1}$. Later work by Green and Ruzsa [16] gave $F(r, K) \leqslant K^2 r^{2K^2-2}$, which was again refined for $r = 2$ to $F(K) \leqslant 2^{O(K^{3/2}\log K)}$ by Sanders [32]. Using compressions, Green and Tao [17] were able to prove $F(K) \leqslant 2^{2K+O(\sqrt{K}\log K)}$. Note that this confirms Conjecture 3.2 for $r = 2$. The best bound so far is due to Konyagin [23], who further improved this method to derive $F(K) \leqslant 2^{2K+O(\log K)}$.

The range of small K has received some attention as well. In the sub-critical range $K < 2$, the exact value of $F(K)$ is known to be $F(K) = K$ for $1 \leqslant K < 7/4$ and $F(K) = \frac{8}{7}K$ for $7/4 \leqslant K < 2$. See [6, 17, 19, 26, 41]. For $K \leqslant 12/5$ we have $F(K) \leqslant (2K-1)/(3K - K^2 - 1)$ and for $12/5 < K < 4$, a recursive formula is available. See [5].

The following simple construction [31] provides a lower bound on $F(K)$.

Example 3.3 (Independent points). Consider the subset

$$A_{[t]} = \{0, e_1, e_2, \ldots, e_t\} \subseteq \mathbb{Z}_2^t.$$

Here, for $t \in \mathbb{N}$ we have

$$F\left(\frac{\binom{t}{2} + t + 1}{t+1}\right) \geqslant \frac{2^t}{t+1},$$

and by monotonicity one can obtain

$$F(K) \geqslant \frac{1}{4K} 2^{2K}(1 - o(1)).$$

3.2. Asymptotics of $F(K)$

We first prove a new upper bound, which coincides with the construction in Example 3.3 for $t \in \mathbb{N}$.

Theorem 3.4. *$F\left(\frac{\binom{t}{2}+t+1}{t+1}\right) \leqslant \frac{2^t}{t+1}$ holds for $2 \leqslant t \in \mathbb{R}$. Consequently,*

$$F(K) \leqslant \frac{1}{2K} 2^{2K}(1 - o(1)).$$

The exponential term 2^{2K} is as in [17, 23], but the polynomial coefficient $1/K$ is new. Thus it re-proves Conjecture 3.2 for $r = 2$ with $C = 2$. This bound and Example 3.3 determine the asymptotics of $F(K)$ up to a factor of 2. In the next section we calculate $F(K)$ exactly, and show that the gap is unavoidable and results from the oscillations in $F(K)$.

Proof. For an affinely generating subset $A \subset G = \mathbb{Z}_2^n$, it is sufficient to prove

$$|A| = \frac{t+1}{2^t}|G| \quad \Rightarrow \quad |A + A| \geqslant \frac{\binom{t}{2} + t + 1}{2^t}|G|, \tag{3.1}$$

where $2 \leqslant t \in \mathbb{R}$. Since both expressions are monotone in t, the theorem follows.

As in [17], the main tool is reduction to compressed sets of some sort. First, since $\langle A \rangle = G$ we can assume that A contains an affine basis for G. But $|A|, |A + A|$ are

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

not affected by invertible affine transformations, so we may assume without loss of generality $E \subseteq A$, where $E = \{0, e_1, e_2, \ldots, e_n\}$ is the standard affine basis of G. Now we assume without loss of generality that A is $\langle\langle E\rangle\rangle$-compressed. Indeed, supposing (3.1) holds for $\langle\langle E\rangle\rangle$-compressed subsets, we proceed to general subsets inducting on $\hbar(A)$. Let $I \subseteq [n]$ be a set such that $E \subseteq C_I(A) \neq A$. By Lemma 2.4, $|C_I(A) + C_I(A)| \leqslant |A + A|$ while $|C_I(A)| = |A|$, so A satisfies (3.1) provided that $C_I(A)$ does. The inductive argument applies, since $\hbar(A) > \hbar(C_I(A))$ by Lemma 2.3(iii).

We continue the proof using the structure of $\langle\langle E\rangle\rangle$-compressed sets. As in Lemma 2.5 let $H \subseteq A$ be a maximal subgroup, $h = \dim H$, $m = \operatorname{codim} H$ and $A_i = A \cap (e_{h+i} + H)$ for $1 \leqslant i \leqslant m$. By Lemma 2.5(vii), $|A| \leqslant (1 + m/2)|H|$, and an upper bound on m is given by

$$\frac{1 + \frac{m}{2}}{2^m} \geqslant \frac{|A|}{|G|},$$

where the case $m = 1$ follows from the assumption $2 \leqslant t$.

Given m, Lemma 2.5(iv) gives a decomposition of A into $m + 1$ parts, and we use it to show that $A + A$ is at least $\sim m/2$ times larger than A. This is shown by the following calculation, where all indices go from 1 to m and all unions are disjoint:

$$\begin{aligned} & A = H \cup \bigcup_i A_i \\ \Rightarrow \quad & A + A = H \cup \bigcup_i (A_i + H) \cup \bigcup_{i<j} (A_i + A_j), \\ & \sum_{i<j} |A_i + A_j| \geqslant \sum_{i<j} \max\big(|A_i|, |A_j|\big) \geqslant \sum_{i<j} \frac{|A_i| + |A_j|}{2} \\ & \qquad = \frac{m-1}{2} \sum_i |A_i| = \frac{m-1}{2}\big(|A| - |H|\big) \\ \Rightarrow \quad & |A + A| \geqslant |H| + m|H| + \frac{m-1}{2}\big(|A| - |H|\big) = \frac{m+3}{2} \cdot \frac{|G|}{2^m} + \frac{m-1}{2}|A|. \end{aligned}$$

The right-hand side is decreasing in m in the real interval where $((m+3)\log 2 - 1)/2^m > |A|/|G|$. This interval includes the range of our interest, which is $(m/2 + 1)/2^m \geqslant |A|/|G| = (t+1)/2^t$, or equivalently $m \leqslant t - 1$. Thus, we obtain a lower bound on $|A + A|$ by evaluating this expression at $t - 1$, namely

$$|A + A| \geqslant \frac{(t-1)+3}{2^{(t-1)+1}}|G| + \frac{(t-1)-1}{2} \cdot \frac{t+1}{2^t}|G| = \frac{\binom{t}{2} + t + 1}{2^t}|G|. \qquad \square$$

3.3. Exact calculation of $F(K)$

Theorem 1.2, which we will shortly prove, provides an explicit formula of $F(K)$. This enables one to rederive the asymptotics of $F(K)$, and to deduce the following corollary.

Corollary 3.5. *Both bounds in the asymptotic inequalities*

$$\frac{1}{4K} 2^{2K}(1 - o(1)) \leqslant F(K) \leqslant \frac{1}{2K} 2^{2K}(1 - o(1))$$

are sharp up to the $o(1)$ *terms.* $\square$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

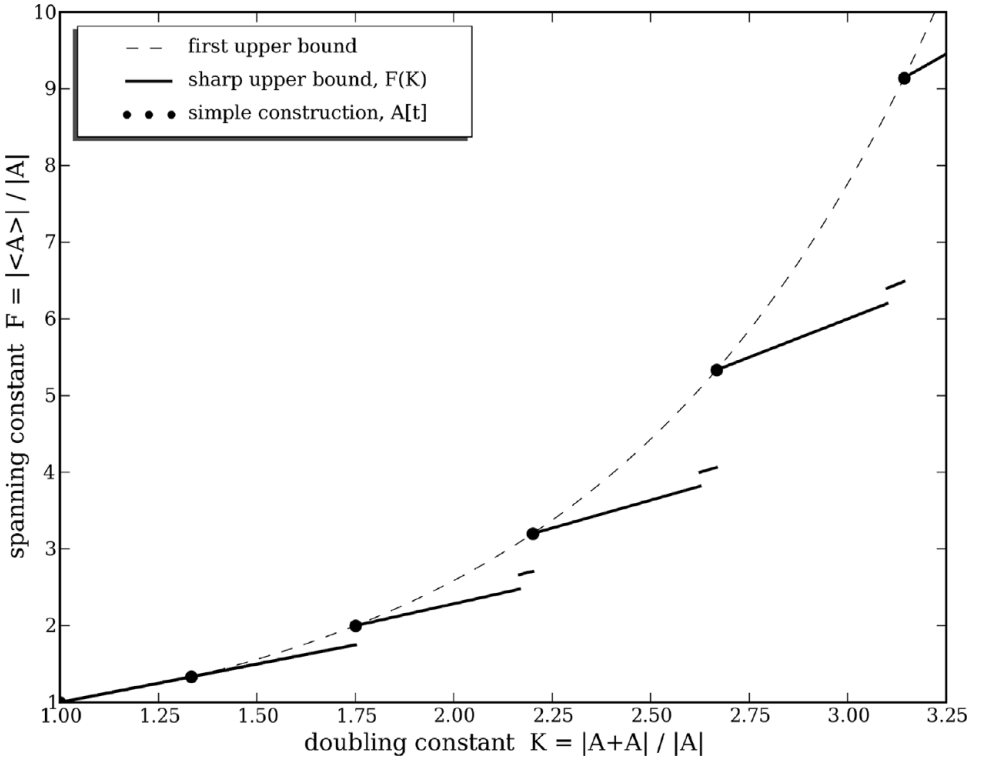

Figure 1. An illustration of $F(K)$.

It also settles the following conjecture of Diao [6].

Corollary 3.6. *$F(K)$ is a piecewise linear function.* □

In order to calculate $F(K)$, it is useful to consider a related function $\tilde{K}(\tilde{F})$, which is defined for rational numbers of the form $\tilde{F} = 2^a/b \geqslant 1$:

$$\tilde{K}(\tilde{F}) = \inf\left\{ \frac{|A+A|}{|A|} \;\middle|\; A \subseteq \mathbb{Z}_2^n,\ n \in \mathbb{N},\ \frac{|\langle A\rangle|}{|A|} = \tilde{F} \right\}.$$

That is, the minimal doubling constant of an affinely generating set of relative size exactly $1/\tilde{F}$. By definition, $F(K) = \sup\{\tilde{F} \mid \tilde{K}(\tilde{F}) \leqslant K\}$. Theorem 3.4 asserts $\tilde{K}(2^t/(t+1)) \geqslant (\binom{t}{2} + t + 1)/(t+1)$ for real $t \geqslant 2$, and by Example 3.3 it is an equality for $t \in \mathbb{N}$. In order to analyse $\tilde{K}(\tilde{F})$, we refine the arguments in the proof of Theorem 3.4, and elaborate on the construction in Example 3.3. This yields a better view of the structure of sets with a small doubling constant. We begin by describing the extended example.

Example 3.7. For non-negative integers s, t such that $s < t$, consider the subset

$$A_{[t,s]} = \{0, e_0, e_1, e_2, \ldots, e_t, e_0 + e_1, e_0 + e_2, \ldots, e_0 + e_{t-s}\} \subseteq \mathbb{Z}_2^{t+1}.$$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

It is not hard to verify that

$$|A_{[t,s]}| = 2(t+1) - s, \quad |A_{[t,s]} + A_{[t,s]}| = 2\left(\binom{t}{2} + t + 1\right) - \binom{s}{2}, \quad |\langle A_{[t,s]}\rangle| = 2^{t+1}.$$

Therefore

$$\tilde{K}\left(\frac{2^t}{t+1-s/2}\right) \leqslant \frac{\binom{t}{2} + t + 1 - \binom{s}{2}/2}{t+1-s/2}.$$

This example provides an upper bound on $\tilde{K}(\tilde{F})$ for a discrete sequence of values. When $s = 0$ it reduces to Example 3.3. However, $\tilde{K}(\tilde{F})$ is not necessarily monotone, so we cannot imitate the conclusion of Example 3.3 and extend the upper bound to general $\tilde{F}$. Still, the following argument does the work.

Lemma 3.8 (Sublinearity of $\tilde{K}(\tilde{F})$). *If $F_1 < F_2$ are in $\tilde{K}$'s domain, then $\frac{\tilde{K}(F_1)}{F_1} \geqslant \frac{\tilde{K}(F_2)}{F_2}$.*

Proof. Let $F_2 = 2^a/b$ for some $a, b \in \mathbb{N}$. Suppose $A_1 \subseteq \mathbb{Z}_2^n$ is an affinely generating set of size $|A_1| = 2^n/F_1$. Let $m \in \mathbb{N}$ be large enough such that $a \leqslant n + m < b2^{n+m-a}$. Consider $A_1' = A_1 \times \mathbb{Z}_2^m$, and note that A_1' affinely generates $\mathbb{Z}_2^{n+m}$ and $|A_1'| = 2^{n+m}/F_1$. Since $F_1 < F_2$ one can take a subset $A_2 \subseteq A_1'$ of cardinality $|A_2| = b2^{n+m-a} = 2^{n+m}/F_2$. Moreover, by m's choice $n + m + 1 \leqslant |A_2|$, so a subset A_2 which affinely generates $\mathbb{Z}_2^{n+m}$ can be chosen. Now from $A_2 + A_2 \subseteq A_1' + A_1' = (A_1 + A_1) \times \mathbb{Z}_2^m$,

$$\frac{|A_1 + A_1|}{|A_1|} \cdot \frac{1}{F_1} = \frac{|A_1 + A_1|}{2^n} = \frac{|A_1' + A_1'|}{2^{n+m}} \geqslant \frac{|A_2 + A_2|}{2^{n+m}} = \frac{|A_2 + A_2|}{|A_2|} \cdot \frac{1}{F_2} \geqslant \frac{\tilde{K}(F_2)}{F_2}.$$

The task is accomplished by taking the infimum over A_1. □

Corollary 3.9 (Superlinearity of $F(K)$). *$\frac{F(K_1)}{K_1} \leqslant \frac{F(K_2)}{K_2}$ for every $1 \leqslant K_1 < K_2$.* □

Example 3.7 and Lemma 3.8 supply an upper bound on $\tilde{K}(\tilde{F})$. The following lemma essentially claims that this bound is sharp.

Lemma 3.10 (Formula for $\tilde{K}(\tilde{F})$). *Let $\tilde{F} \geqslant 1$ be of the form $2^a/b$ where $a, b \in \mathbb{N}$, and let $s < t$ be the unique pair of non-negative integers for which*

$$\frac{2^t}{t+1-s/2} \leqslant \tilde{F} < \frac{2^t}{t+1-(s+1)/2}.$$

Then

$$\tilde{K}(\tilde{F}) = \frac{\binom{t}{2} + t + 1 - \frac{1}{2}\binom{s}{2}}{2^t} \cdot \tilde{F}.$$

Since the function $F(K)$ is basically the inverse of $\tilde{K}(\tilde{F})$, Theorem 1.2 is a direct consequence of Lemma 3.10. Indeed, Figure 1 is obtained by transposing the graph in Figure 2, and taking the maximum wherever the result is multivalued. We omit further details.

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

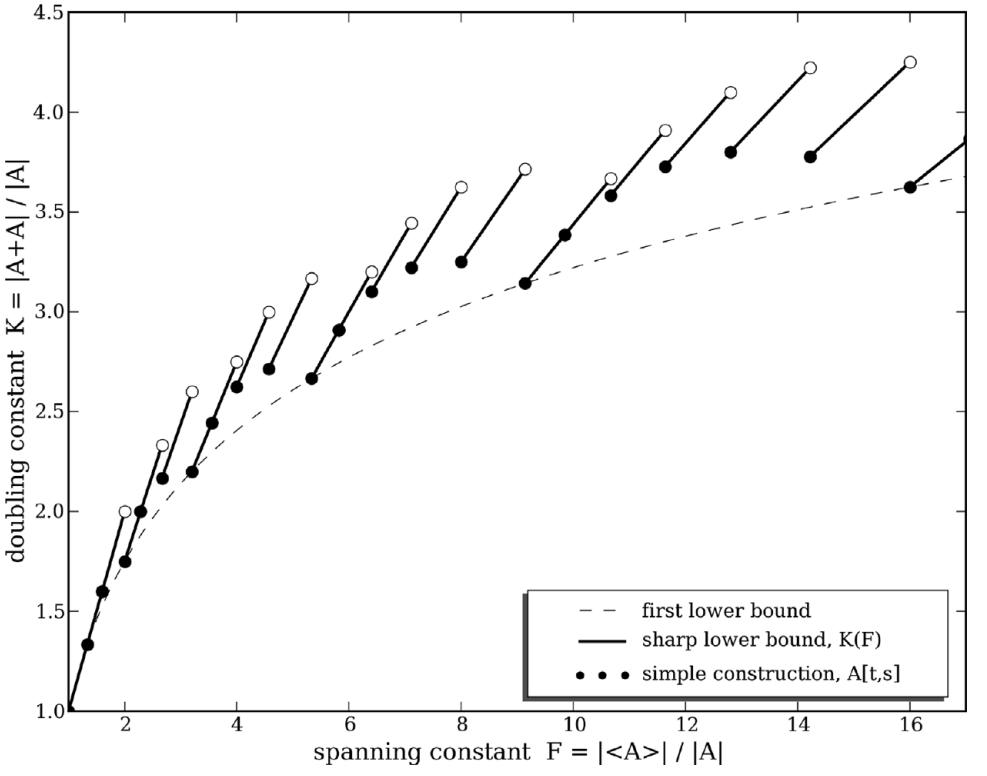

Figure 2. An illustration of $\tilde{K}(\tilde{F})$.

One can notice that $\tilde{K}(\tilde{F})$ has a more complex structure than $F(K)$. Since Theorem 1.2 employs the information in Lemma 3.10 only partially, there may be a quicker way of calculating $F(K)$. Nevertheless, we feel that the detailed description of $\tilde{K}(\tilde{F})$ is interesting in its own right, and may shed light on the non-trivial form of $F(K)$.

The proof of Lemma 3.10 pursues the analysis in Theorem 3.4's proof, involving more reduction steps which preserve $|A|$ without increasing $|A+A|$. Through these reductions the structure of A becomes similar to Example 3.7, so that its doubling constant can be calculated explicitly. We start with two reductions which can be formulated separately in terms of integer partitions. All of the following will be motivated and applied later, in the proof of the lemma.

A non-increasing sequence of positive integers $a_1 \cdots a_m$ is an integer *partition* of $a = \sum_i a_i$ into m parts, and, for short, an *m-partition* of a. Recall the Hopf–Stiefel function $a \circ b$ from Section 2. We are interested in the minimum of $\sum_{1 \leqslant i < j \leqslant m} a_i \circ a_j$ over all m-partitions of a.

A partition $a_1 \cdots a_m$ of a is called *compressed* if $a_i + a_j > 2^k \Rightarrow a_i \geqslant 2^k$ for each k and $i < j$. It will be implicit in the proof of Lemma 3.10 that at least one of the partitions that minimize $\sum_{i<j} a_i \circ a_j$ is compressed. Here we shall restrict the discussion to compressed partitions.

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

A partition is called *quasi-dyadic* if $a_1 \cdots a_{m-1}$ are powers of 2. No requirement is made on a_m. Note that a quasi-dyadic partition is always compressed. Our first reduction basically asserts that the minimum of $\sum_{i<j} a_i \circ a_j$ is attained by a quasi-dyadic partition.

Lemma 3.11. *A compressed m-partition of a that minimizes $\sum_{i<j} a_i \circ a_j$ is quasi-dyadic.*

Proof. Otherwise, consider the smallest $1 \leqslant i < m$ for which a_i is not a power of 2, say $2^k < a_i < 2^{k+1}$. Since the partition is compressed, $a_i + a_{i+1} \leqslant 2^{k+1}$. We 'transfer mass' from a_i to a_{i+1}. Replace a_i by $a'_i = 2^k$, and replace a_{i+1} by $a'_{i+1} = a_{i+1} + a_i - 2^k$. Note that $a'_i \geqslant a'_{i+1}$ and monotonicity is preserved.

How does this move affect $\sum_{i<j} a_i \circ a_j$? By the choice of i, for $j < i$ we have $a_j = 2^l$, where $l > k$ as a partition is non-increasing. Thus the terms involving a_j are unchanged:

$$a_j \circ a_i + a_j \circ a_{i+1} = 2^l + 2^l = a_j \circ a'_i + a_j \circ a'_{i+1}.$$

For $j > i+1$ we know $a_j \leqslant a_{i+1} < 2^k$. By the recursive definition of the Hopf–Stiefel function, and the sub-distributive law,

$$\begin{aligned} a_j \circ a_i + a_j \circ a_{i+1} &= 2^k + a_j \circ (a_i - 2^k) + a_j \circ a_{i+1} \\ &\geqslant 2^k + a_j \circ (a_i - 2^k + a_{i+1}) = a_j \circ a'_i + a_j \circ a'_{i+1}. \end{aligned}$$

Finally, again by the recursive definition the mixed term becomes strictly smaller:

$$a_i \circ a_{i+1} = 2^k + (a_i - 2^k) \circ a_{i+1} > 2^k = a'_i \circ a'_{i+1}.$$

The combination of the last three calculations yields that the sum $\sum_{i<j} a_i \circ a_j$ can be made smaller by changing the partition, in contradiction to the minimality assumption. □

Since $2^k \circ a = 2^k$ for $a \leqslant 2^k$, in the quasi-dyadic case the summation can be simplified:

$$\sum_{1 \leqslant i < j \leqslant m} a_i \circ a_j = \sum_{1 \leqslant i < j \leqslant m} \max(a_i, a_j) = \sum_{1 \leqslant i < j \leqslant m} a_i = \sum_{i=1}^{m} (m-i) \cdot a_i.$$

It is natural to conjecture that the minimum is obtained when $a_1 \cdots a_m$ are 'almost' equal. A quasi-dyadic m-partition of a is *quasi-fair* if, for some $k \in \mathbb{N}$, $a_i \in \{2^k, 2^{k-1}\}$ for each $1 \leqslant i \leqslant m-1$. For example, $4+4+2+2+2+1$ and $4+4+4+3$ and $8+4+3$ are some quasi-fair partitions of 15. The following properties of quasi-fair partitions are easily verified.

(1) In the above definition one can choose

$$k = \lceil \log_2(a/m) \rceil,$$

and then exactly $a_1 \cdots a_j$ exceed 2^{k-1}, where

$$j = \lceil a/2^{k-1} \rceil - m.$$

(2) For every two positive integers $m \leqslant a$, there exists a unique quasi-fair m-partition of a.

(3) If $a_1 \cdots a_m$ and $a'_1 \cdots a'_m$ are the quasi-fair m-partitions of $a \leqslant a'$, then $a_i \leqslant a'_i$ for all i.

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

(4) A sub-partition (in the sense of a sub-sequence) of a quasi-fair partition is quasi-fair.

Now we are ready to state the second reduction.

Lemma 3.12. *The minimum of $\sum_{i<j} a_i \circ a_j$ over all quasi-dyadic m-partitions of a is obtained only by the quasi-fair one.*

Proof. This lemma can be verified by induction on m. For a partition that minimizes the sum, it is enough to show $a_1 = 2^k$ for $k = \lceil \log_2(a/m) \rceil$. By the induction hypothesis $a_2 \cdots a_m$ are quasi-fair and thus constitute the unique quasi-fair sub-partition we are looking for. By the monotonicity property applied on $a_2 \cdots a_m$, for a competing sequence $a_1' \cdots a_m'$ with $a_1' > a_1$, necessarily $a_i' \leqslant a_i$ for $i \geqslant 2$, and consequently

$$\begin{aligned}\sum_{i=1}^{m}(m-i)\cdot a_i &= \sum_{i=1}^{m}(m-i)\cdot a_i + (m-1)\left(\sum_{i=1}^{m} a_i' - \sum_{i=1}^{m} a_i\right)\\ &= \sum_{i=1}^{m}(m-i)\cdot a_i' - \sum_{i=2}^{m}(i-1)(a_i - a_i') < \sum_{i=1}^{m}(m-i)\cdot a_i'.\end{aligned}$$

□

With these reductions in hand, we can complete the calculation of $\tilde{K}(\tilde{F})$.

Proof of Lemma 3.10. Let $\langle A \rangle = G = \mathbb{Z}_2^n$. Lemma 3.10 is proved by showing the following lower bound on $|A + A|$, which is reached by Example 3.7 and Lemma 3.8:

$$\frac{t+1-(s+1)/2}{2^t} < \frac{|A|}{|G|} \leqslant \frac{t+1-s/2}{2^t} \quad \Rightarrow \quad \frac{|A+A|}{|G|} \geqslant \frac{\binom{t}{2} + t + 1 - \frac{1}{2}\binom{s}{2}}{2^t}. \tag{3.2}$$

If $|A| > \frac{1}{2}|G|$, then by the pigeonhole principle $A + A = G$, as required in the cases $t = 1, 2$. Hence we may assume $|A| \leqslant \frac{1}{2}|G|$ and $t \geqslant 3$.

We start as in Theorem 3.4. We first assume without loss of generality that A is $\langle\langle E \rangle\rangle$-compressed and therefore, by Lemma 2.5, has the following properties.

- There exists a subgroup $H = \langle 0, e_1, \ldots, e_h \rangle$ such that $A = H \cup A_1 \cup A_2 \cup \cdots \cup A_m$, where $A_i = A \cap (e_{h+i} + H)$ and $m = \operatorname{codim} H$.
- $1/2^m < |A|/|G| \leqslant (1 + \frac{m}{2})/2^m$. By the assumptions $\frac{t+2}{2^{t+1}} < |A|/|G| \leqslant \frac{1}{2}$, we can write $1 < m < t$.
- Each A_i is a lexicographic initial segment of $e_i + H$. Therefore A is uniquely determined by the sequence $a_1, \ldots, a_m$, where $a_i = |A_i|$. Note that $0 < a_i < 2^h$.
- By shift-minimality $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_m$. In other words, $a_1, \ldots, a_m$ is a partition of $a = |A| - |G|/2^m$.

As in Theorem 3.4, we use these properties to write $A + A$ as a disjoint union of its intersections with H-cosets, which are of three forms: H, $H + A_i$ and $A_i + A_j$. Since the A_i are initial segments of their cosets, the sumsets of the third form can be expressed via the Hopf–Stiefel function:

$$|A + A| = |H| + m|H| + \sum_{1 \leqslant i < j \leqslant m} |A_i + A_j| = \frac{m+1}{2^m} \cdot |G| + \sum_{1 \leqslant i < j \leqslant m} a_i \circ a_j.$$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

This equation makes it interesting to find partitions $a_1 \cdots a_m$ of a that minimize $\sum_{i<j} a_i \circ a_j$.

We next show that the partition $a_1 \cdots a_m$ is compressed. For $i < j$ and $1 \leqslant k \leqslant h$ we exclude the case where $a_i < 2^k < a_i + a_j$ by the assumption that A is already $\langle\langle E \rangle\rangle$-compressed. For $I = \{1, 2, \ldots, k, h+i, h+j\}$, let us examine the set $C_I(A)$. A_i is replaced by an initial segment of $e_{h+i} + H$ of size 2^k, and A_j is replaced by an initial segment of $e_{h+j} + H$ of size $a_i + a_j - 2^k$, which is not empty by assumption. In other words, a_i becomes 2^k, and $E \subseteq A$ is preserved.

By Lemmas 3.11–3.12, if $|A + A|$ is minimal then $a_1 \cdots a_m$ is the quasi-fair quasi-dyadic m-partition of $a = |A| - |G|/2^m$. In this situation

$$|A + A| - \frac{m+1}{2^m} \cdot |G| = \sum_{i=1}^{j} (m-i) \cdot 2^k + \sum_{i=j+1}^{m} (m-i) \cdot 2^{k-1} = \left[\binom{m}{2} - \frac{1}{2} \binom{m-j}{2} \right] 2^k$$

for $0 \leqslant j \leqslant m$ and $0 < k < (\dim G - m)$ such that

$$\frac{m+j-1}{2} \cdot 2^k < |A| - \frac{|G|}{2^m} \leqslant \frac{m+j}{2} \cdot 2^k.$$

Remark. Note that in the cases $j = m-1$ or $j = m$ we can choose $j' = 0$ and $k' = k+1$ as well. We could avoid this freedom of choice by not permitting $j = 0$, but since it does not affect the resulting $|A + A|$, we allow both ways.

All that remains now is to show that, as in Theorem 3.4, to minimize $|A + A|$ we should make m as large as possible, *i.e.*, $m = t - 1$. The proof is by induction on $t - m$.

- Suppose $m = t - 1$. We check that (3.2) holds:

  $$\frac{2m - (s+1)}{2} \cdot \frac{|G|}{2^{m+1}} < |A| - \frac{|G|}{2^m} \leqslant \frac{2m - s}{2} \cdot \frac{|G|}{2^{m+1}}.$$

  Denote $k = \dim G - m - 1$ and $j = m - s$, and observe that $0 \leqslant j \leqslant m$. Then the above expression for the minimal $|A + A|$ becomes

  $$|A + A| = \frac{m+1}{2^m} \cdot |G| + \left[\binom{m}{2} - \frac{1}{2} \binom{m - \binom{m-s}{}}{2} \right] \cdot \frac{|G|}{2^{m+1}} = \frac{\binom{t}{2} + t + 1 - \frac{1}{2}\binom{s}{2}}{2^t} \cdot |G|.$$

- Suppose $m < t - 1$. The above discussion yields a compressed set A, such that $a_1 \cdots a_m$ is the quasi-fair quasi-dyadic partition of $|A| - |H|$, and $|A + A|$ is minimal given m, and equals

  $$|A + A| = \frac{m+1}{2^m} |G| + \sum_{1 \leqslant i < j \leqslant m} a_i \circ a_j.$$

  We show that increasing m makes $|A + A|$ smaller. Denote by A', H' and $a'_1 \cdots a'_{m+1}$ the corresponding set, subgroup and partition for $m' = m + 1$. Similarly,

  $$|A' + A'| = \frac{m+2}{2^{m+1}} |G| + \sum_{1 \leqslant i < j \leqslant m+1} a'_i \circ a'_j.$$

  Now define $a_0 = |H'| = |H|/2 = |G|/2^{m+1}$. Since $a_1 \cdots a_m$ is a quasi-dyadic m-partition for $m > 1$ and $a_1 < |H|$, necessarily $a_0 \geqslant a_i$ and $a_0 \circ a_i = a_0$ for all $1 \leqslant i \leqslant m$. Hence,

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

for the quasi-dyadic $(m+1)$-partition $a_0 \cdots a_m$,

$$|A+A| = \left(\frac{m+1}{2^m} - \frac{m}{2^{m+1}}\right)|G| + m \cdot a_0 + \sum_{1 \leqslant i < j \leqslant m} a_i \circ a_j = \frac{m+2}{2^{m+1}}|G| + \sum_{0 \leqslant i < j \leqslant m} a_i \circ a_j.$$

But by Lemma 3.12, the partition $a_1' \cdots a_{m+1}'$ gives the minimal value for this expression. Moreover, since $a_0 = |H'| > a_1'$, these partitions differ and $|A'+A'| < |A+A|$. □

Remark. An examination of the proof reveals two kinds of reduction steps. Either A is compressed without changing $\langle A \rangle$, or we find a set A' where $|A'| = |A|$ and $|A'+A'|$ is substantially smaller than $|A+A|$. Hence, the proof actually provides a characterization of the extremal case, up to compressions that preserve $\langle A \rangle$ and $|A+A|$.

4. Addition of two different sets

What is the smallest possible cardinality of $A+B$ if $A, B \subseteq G = \mathbb{Z}_2^n$ are two affinely spanning subsets of given cardinalities? In this section we prove Theorem 1.1, which gives an essentially complete answer. In addition we establish a new isoperimetric inequality, which is used in the proof. But first, we make some remarks concerning the theorem.

Remarks on Theorem 1.1.
(1) *Tightness.* Consider $(|A|/|G|, |B|/|G|, |A+B|/|G|)$ as a point in $[0,1]^3$. The Hamming balls construction shows that the bound goes through the points of the form

$$\left(\frac{1+t}{2^t}, \quad \frac{1+t+\cdots+\binom{t}{k}}{2^t}, \quad \frac{1+t+\cdots+\binom{t}{k+1}}{2^t}\right), \quad 0 \leqslant k < t.$$

An inspection of Figure 3 shows that all points properly inside their convex hull are strictly below the bound, and hence cannot be realized by such sets. In other words, further improvements of the bound will be local in nature.

(2) The formulation of the theorem apparently breaks the symmetry and does not require $\langle B \rangle = G$. Still, there is an asymmetry in the result as well, and the theorem is of interest mostly when $|A| \leqslant |B|$. See also the remark after the proof.

(3) The only assumption on t which the proof uses is $(t+2)/2^{t+1} < |A|/|G|$. The statement can, therefore, be applied as well with t larger than in the theorem. As Figure 3 shows, the resulting bound would be weaker, but may still be useful in certain contexts.

Theorem 1.1 implies that a large enough number of large enough affinely generating sets must add up to the whole group.

Corollary 4.1. *Suppose that* $\langle A_1 \rangle = \cdots = \langle A_m \rangle = G = \mathbb{Z}_2^n$ *with* $|A_i|/|G| > (m+2)/2^{m+1}$ *for all* i. *Then* $A_1 + A_2 + \cdots + A_m = G$.

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

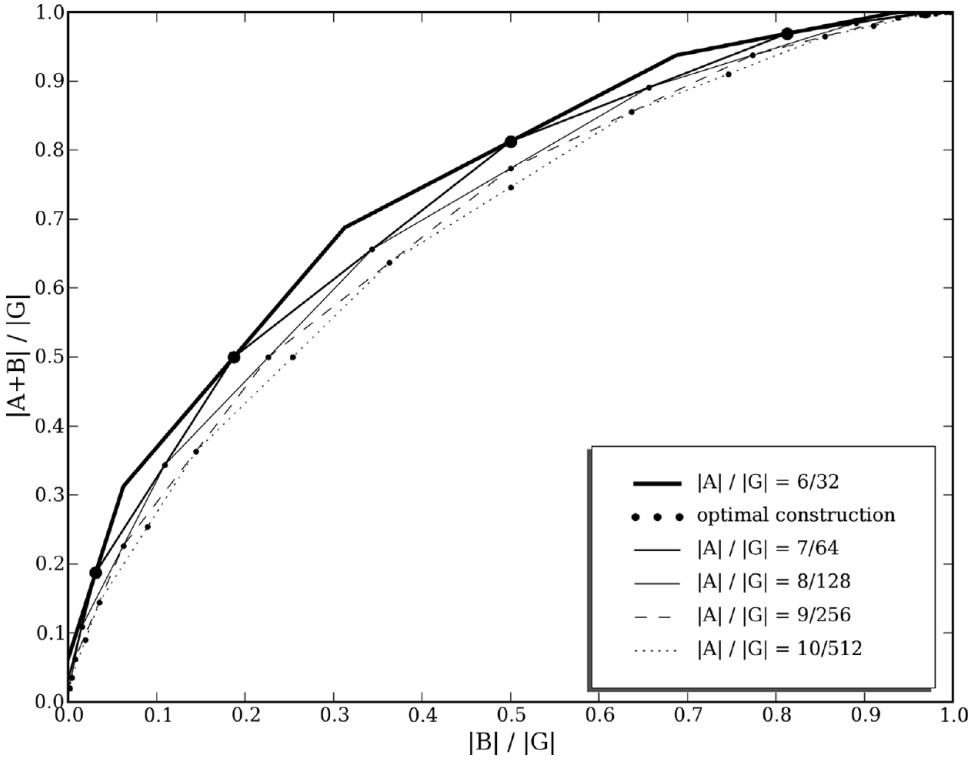

Figure 3. An illustration of the lower bound.

Proof. We repeatedly apply the theorem with $A = A_i$ and $B = A_1 + \cdots + A_{i-1}$ for all $1 \leqslant i \leqslant m$, to conclude

$$\frac{|A_1 + A_2 + \cdots + A_i|}{|G|} > \frac{\binom{m+1}{0} + \binom{m+1}{1} + \cdots + \binom{m+1}{i}}{2^{m+1}}.$$

Indeed, in view of remark (3) above and the assumption on the cardinalities, we may choose $t = m + 1$, and then $k = i - 1$ and $w > 0$ by the induction hypothesis. Since $|A_1 + A_2 + \cdots + A_{m-1}| + |A_m| > |G|$, the proof is completed by the pigeonhole principle, $|A| + |B| > |G| \Rightarrow A + B = G$. □

The special case of Corollary 4.1 where all A_i are identical is due to Lev [25], following a conjecture of Zemor [40]. Taking $A_i = D_1^{m+1} \times \mathbb{Z}_2^{n-m-1}$ for each i shows that the assumption on the cardinalities is sharp.

4.1. An isoperimetric inequality

We are inspired by Frankl's short inductive proof [13] of Harper's theorem [18].

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

Theorem 4.2 (Harper's inequality). *Suppose $A \subseteq \mathbb{Z}_2^n$. If, for $1 \leqslant k \leqslant n$ integer and $0 \leqslant p \leqslant 1$ real,*

$$|A| = \binom{n}{n} + \binom{n}{n-1} + \cdots + \binom{n}{k+1} + p\binom{n}{k},$$

then

$$|A + D_1^n| \geqslant \binom{n}{n} + \binom{n}{n-1} + \cdots + \binom{n}{k} + p\binom{n}{k-1}.$$

In simple terms this theorem says that Hamming balls solve the vertex-isoperimetric problem in the hypercube. However, it also deals with sets of cardinalities strictly between $|D_{k-1}^n|$ and $|D_k^n|$, to varying degrees depending on the version of the theorem. A stronger version would replace the last summand of each expression with $\binom{x}{k}$ and $\binom{x}{k-1}$ respectively, where $x \in [k, n]$ is real. The optimal formulation due to Katona [22] and Kruskal [24] is stated in terms of the k-cascade representations $\binom{a_k}{k} + \binom{a_{k-1}}{k-1} + \cdots$ and $\binom{a_k}{k-1} + \binom{a_{k-1}}{k-2} + \cdots$ respectively. Frankl's method yields all three formulations.

Frankl's proof employs several useful operators on set-systems. As usual, we freely move between the set-theoretic terminology of $2^{[n]}$ and the algebraic language of $\mathbb{Z}_2^n$. The push-down operator T_i and the shift operator S_{ij} have already appeared in Section 2. The upper and the lower *shadow* operators act on a set-system $\mathcal{F} \subseteq 2^{[n]}$ by

$$\delta\mathcal{F} = \{J \cup \{i\} \mid J \in \mathcal{F}, i \notin J\},$$
$$\partial\mathcal{F} = \{J \setminus \{i\} \mid J \in \mathcal{F}, i \in J\},$$

respectively. For downsets, the notion of the shadow is close to that of the neighbourhood in the theorem. If $C \subseteq \mathbb{Z}_2^n$ is a non-empty downset, then $C + D_1^n = \delta C \cup \{0\}$. Note that always $0 \notin \delta A$. Another useful operation on set-systems is *classification* by n, denoted by

$$\mathcal{F}^- = \{J \mid J \in \mathcal{F}, n \notin J\},$$
$$\mathcal{F}^+ = \{J \setminus \{n\} \mid J \in \mathcal{F}, n \in J\}.$$

When $A \subseteq \mathbb{Z}_2^n$, we regard A^+ and A^- as subsets of $\mathbb{Z}_2^{n-1}$.

Following Frankl [13], we proceed with two lemmas regarding properties of shifts and shadows.

Lemma 4.3. *Suppose $C \subseteq \mathbb{Z}_2^n$ is a shift-minimal downset:*

(i) $\delta(C^+) \subseteq (\delta C)^+ = C^-$ *with equality if and only if* $C = \varnothing$,
(ii) $\delta(C^-) = (\delta C)^-$.

Proof. Examine the effect of the operators on the representation of some $x \in C$ with the standard basis $e_1, \ldots, e_n$.

In both $\delta(C^+)$ and $(\delta C)^+$, some e_i is added and e_n is removed. However, in $\delta(C^+)$ certainly $i \neq n$ since C^+ lives in $\mathbb{Z}_2^{n-1}$, while in $(\delta C)^+$ it is possible that $i = n$. Hence $\delta(C^+) \subseteq (\delta C)^+$. By shift-minimality C is closed under these swaps, thus $(\delta C)^+ \subseteq C^-$.

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

Moreover, every element of C^- is obtained by adding e_n and then deleting it, so there is equality. However, $\delta(C^+)$ is strictly smaller since $0 \in C^- \setminus \delta(C^+)$ unless C is empty.

For $\delta(C^-) = (\delta C)^-$, note that both sets consist of elements of the form $x + e_i$ for $x \in C$ and $i < n$, where e_i and e_n do not appear in x's standard representation. □

The following lemma is well known: see, *e.g.*, [11, 21]. Here we prove it as a special case of the compression machinery.

Lemma 4.4. *For all $A \subseteq \mathbb{Z}_2^n$ and $1 \leqslant i, j \leqslant n$ such that $i \neq j$:*

(i) $\delta(S_{ij}A) \subseteq S_{ij}(\delta A)$,
(ii) $\partial(S_{ij}A) \subseteq S_{ij}(\partial A)$.

Proof. By passing from A to $\sum_i e_i - A$, it is enough to prove only one of the inclusions. Denote $A = \bigcup_{k=0}^n A_k$ where $A_k = A \cap (D_k^n \setminus D_{k-1}^n)$. Note that we can work with each A_k separately. One can write

$$\delta(S_{ij}A_k) = (D_1^n + C_{ij}(A_k \cup D_{k-1}^n)) \setminus D_k^n$$

and

$$S_{ij}(\delta A_k) = C_{ij}(D_1^n + (A_k \cup D_{k-1}^n)) \setminus D_k^n,$$

yielding our claim by Lemma 2.4, since $D_1^n + C_{ij}(B) = C_{ij}(D_1^n) + C_{ij}(B) \subseteq C_{ij}(D_1^n + B)$. □

Our isoperimetric inequality concerns a family of non-empty downsets $C_1 \cdots C_l \subseteq \mathbb{Z}_2^n$, rather than a single one. For the volume and the shadow we take the average quantities, denoted by

$$\mathrm{E}[C] = \frac{1}{l}\sum_{m=1}^{l} |C_m|, \quad \mathrm{E}[\delta C] = \frac{1}{l}\sum_{m=1}^{l} |\delta(C_m)|.$$

It is hard to make a meaningful statement about these average quantities without limiting the downsets somehow. To see this, consider what happens when each C_m is either full or empty. We limit the variability of the downsets by assuming the *antichain condition*. Namely, we require that for each i and j, $C_i \setminus C_j$ is an antichain with respect to set-systems inclusion, or equivalently $C_j \supseteq \partial C_i$.

Proposition 4.5. *Suppose $C_1 \cdots C_l \subseteq \mathbb{Z}_2^n$ is a family of downsets which satisfies the antichain condition. If*

$$\mathrm{E}[C] = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k-1} + p\binom{n}{k}$$

for some integer $k \geqslant 0$ and real number $0 \leqslant p < 1$, then

$$\mathrm{E}[\delta C] \geqslant \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k} + p\binom{n}{k+1}.$$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

Since for non-empty downsets $C + D_1^n = \{0\} \cup \delta C$, the corresponding inequality in the language of neighbourhoods is as follows.

Corollary 4.6. *In the setting of Proposition 4.5, if $C_1 \cdots C_l$ are non-empty then*

$$\mathrm{E}[C + D_1^n] \geqslant \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k} + p\binom{n}{k+1}.$$

Proof of Proposition 4.5. We may assume that the downsets are shift-minimal. Indeed, for each downset C_m clearly $S_{ij}C_m$ is a downset of the same size, while $|\delta(S_{ij}C_m)| \leqslant |S_{ij}(\delta C_m)| = |\delta C_m|$ by Lemma 4.4. If $C_{m'} \setminus C_m$ is an antichain, then $C_m \supseteq \partial C_{m'}$. Thus, by Lemma 4.4 again, $S_{ij}C_m \supseteq S_{ij}(\partial C_{m'}) \supseteq \partial(S_{ij}C_{m'})$, and hence $S_{ij}C_{m'} \setminus S_{ij}C_m$ is an antichain as well. In conclusion, $S_{ij}C_1 \cdots S_{ij}C_l$ satisfy the antichain condition, $\mathrm{E}[C] = \mathrm{E}[S_{ij}C]$ and $\mathrm{E}[\delta C] \geqslant \mathrm{E}[\delta(S_{ij}C)]$. After a finite sequence of shifts the downsets are all shift-minimal, since for a proper shift

$$\sum_m \hbar(S_{ij}C_m) < \sum_m \hbar(C_m).$$

The case $k = 0$ is established separately. Note that in this case $\mathrm{E}[C] < 1$, hence $C_m = \varnothing$ for some m. Actually, this is a sufficient condition for $k = 0$, because all other downsets are either $\varnothing$ or $\{0\}$ by the antichain condition. Since $\delta\{0\} = \{e_1, \ldots, e_n\}$, clearly $\mathrm{E}[\delta C] = n \cdot \mathrm{E}[C]$ as required.

Following Frankl, we proceed by induction on n. By convention $\binom{n}{k} = 0$ for $n < k$. Thus, for $n = 0$ the lemma is vacuously satisfied by $\mathrm{E}[\delta C] \geqslant 0$.

For positive k and n, we employ the induction hypothesis on the families $C_1^- \cdots C_l^-$ and $C_1^+ \cdots C_l^+$ in $\mathbb{Z}_2^{n-1}$. It is easily checked that given a downset C_m, the sets C_m^+ and C_m^- are downsets as well. In addition, if $C_{m'} \setminus C_m$ is an antichain, then so are its two parts, $C_{m'}^- \setminus C_m^-$ and $C_{m'}^+ \setminus C_m^+$, hence the new families satisfy the antichain condition.

By the induction hypothesis on $C_1^+ \cdots C_l^+ \subseteq \mathbb{Z}_2^{n-1}$, at least one of the following must hold:

$$\mathrm{E}[C^+] < \binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{k-2} + p\binom{n-1}{k-1},$$
$$\mathrm{E}[\delta(C^+)] \geqslant \binom{n-1}{1} + \binom{n-1}{2} + \cdots + \binom{n-1}{k-1} + p\binom{n-1}{k}.$$

Use $\mathrm{E}[C^-] = \mathrm{E}[C] - \mathrm{E}[C^+]$ and Pascal's rule in the first case, or $\mathrm{E}[C^-] \geqslant 1 + \mathrm{E}[\delta(C^+)]$ by Lemma 4.3(i) in the second one, to deduce

$$\mathrm{E}[C^-] \geqslant \binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{k-1} + p\binom{n-1}{k}.$$

Note that since $k > 0$ each C_m is non-empty, so there is proper inclusion in the lemma, which yields the extra 1 in the calculation. By the induction hypothesis on $C_1^- \cdots C_l^- \subseteq \mathbb{Z}_2^{n-1}$,

$$\mathrm{E}[\delta(C^-)] \geqslant \binom{n-1}{1} + \binom{n-1}{2} + \cdots + \binom{n-1}{k} + p\binom{n-1}{k+1}.$$

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

By Lemma 4.3, $\mathrm{E}[\delta C] = \mathrm{E}[(\delta C)^-] + \mathrm{E}[(\delta C)^+] = \mathrm{E}[\delta(C^-)] + \mathrm{E}[C^-]$. Hence, by Pascal's rule,

$$\mathrm{E}[\delta C] \geqslant \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k} + p\binom{n}{k+1}. \qquad \square$$

4.2. Proof of the lower bound

Proof of Theorem 1.1. The general idea is similar to the case $A + A$ discussed in the previous section. By applying various compressions, the sets A and B acquire certain structural properties. These, in turn, allow us to derive estimates on the cardinality of $A + B$.

Lemma 2.4 asserts that compressions do not increase sumsets: $|C_I(A) + C_I(B)| \leqslant |A + B|$ holds while $|C_I(A)| = |A|$ and $|C_I(B)| = |B|$. Thus, in the search for a lower bound for $|A + B|$, one can first apply a compression C_I on A and B simultaneously. Since $\langle A\rangle = G$, we may suppose $E = \{0, e_1, e_2, \ldots, e_n\} \subseteq A$ and restrict ourselves only to compressions that preserve the inclusion $E \subseteq A$. By Lemma 2.3(iii), if a compression C_I changes either A or B, then $\hbar(A) + \hbar(B)$ strictly decreases. It follows that every sequence of such compressions must terminate. In conclusion, we can assume that both A and B are invariant under these compressions, or, for short, $\langle\langle E \subseteq A\rangle\rangle$-*compressed*. This implies that B is I-compressed for every $I \subseteq [n]$ such that A is I-compressed.

Lemma 2.5 provides a description of A under this assumption. In particular, $H \subseteq A \subseteq H + E$ for some subgroup $H = \langle 0, e_1, \ldots, e_h\rangle$. We next derive some structural properties of B.

Lemma 4.7. *Suppose $A, B \subseteq G = \mathbb{Z}_2^n$ are $\langle\langle E \subseteq A\rangle\rangle$-compressed. Let $H \subseteq A$ be as in Lemma 2.5. Consider $G/H \cong \mathbb{Z}_2^m$, where $m = n - h = \operatorname{codim} H$, with the basis $\{e_{h+1} + H, \ldots, e_{h+m} + H\}$ and the partial order of the corresponding set-system. For $1 \leqslant j \leqslant |H|$ let*

$$C_j = \Big\{H' \in G/H \mid |B \cap H'| \geqslant j\Big\}.$$

Then $C_1 \cdots C_{|H|}$ are downsets, and satisfy the antichain condition.

Proof. By Lemma 2.5(iii), A is $\{1, \ldots, h, h + i\}$-compressed for $1 \leqslant i \leqslant m$, and therefore so is B.

Let $H' \prec H''$ be adjacent H-cosets in the partial order. $H'' = e_{h+i} + H'$ for some $1 \leqslant i \leqslant m$. Since B is $\{1, \ldots, h, h + i\}$-compressed, $B \cap (H' \cup H'')$ must be an initial segment of $H' \cup H''$. Note that all H' elements are lexicographically smaller than those of H''. Consequently, if $B \cap H'' \neq \varnothing$ then necessarily $H' \subseteq B$. In other words, $H'' \in C_1 \Rightarrow H' \in C_{|H|}$ for each such pair.

In particular, C_j is a downset because $H'' \in C_j \subseteq C_1 \Rightarrow H' \in C_{|H|} \subseteq C_j$, and $C_j \setminus C_k$ is an antichain since $C_j \setminus C_k \subseteq C_1 \setminus C_{|H|} \not\supseteq \{H', H''\}$. $\square$

We can now conclude the proof of Theorem 1.1 in the following three steps.

(1) We use the structure of the compressed sets to find new expressions for the cardinalities of B and $A + B$. Let $C_1 \cdots C_{|H|}$ be as in Lemma 4.7. By interchanging the order of

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

summation,

$$|B| = \sum_{H' \in G/H} |B \cap H'| = \sum_{H' \in G/H} \#\Big\{ j \in \mathbb{N} \mid |B \cap H'| \geqslant j \Big\} = \sum_{j=1}^{|H|} |C_j|.$$

We estimate $|A+B|$ in a similar fashion. For $1 \leqslant j \leqslant |H|$, suppose $H'' \in \delta(C_j) \cup \{H\}$. We show that $A+B$ intersects H'' in at least j elements.

- If $H'' = H$, use $H \subseteq A$ and $0 \in B \neq \varnothing$, to obtain $|(A+B) \cap H''| \geqslant |(H+0) \cap H| \geqslant j$.
- Otherwise $H'' = e_{h+i} + H'$ for some $H' \in C_j$ and $1 \leqslant i \leqslant m = \operatorname{codim} H$. Since $e_{h+i} \in E \subseteq A$, clearly $|(A+B) \cap H''| \geqslant |(e_{h+i}+B) \cap (e_{h+i}+H')| = |B \cap H'| \geqslant j$.

Consequently,

$$|A+B| = \sum_{H'' \in G/H} \#\Big\{ j \in \mathbb{N} \mid |(A+B) \cap H''| \geqslant j \Big\} \geqslant \sum_{j=1}^{|H|} |\delta(C_j) \cup \{H\}|.$$

(2) We use the isoperimetric inequality in order to obtain a lower bound on $|A+B|$ given m and $|B|$. Let $0 \leqslant k \leqslant m$ and $w \in [-1,1]$ be such that

$$|B| = \frac{\binom{m+1}{0} + \binom{m+1}{1} + \cdots + \binom{m+1}{k} + w\binom{m}{k}}{2^{m+1}} \cdot |G|.$$

We substitute $|B| = \sum |C_j|$ in the left-hand side, apply Pascal's rule to $\binom{m+1}{1} \cdots \binom{m+1}{k}$ on the right-hand side, and divide both by $|H| = |G|/2^m$, to obtain

$$\mathrm{E}[C] = \frac{1}{|H|} \sum_{j=1}^{|H|} |C_j| = \frac{|B|}{|H|} = \binom{m}{0} + \binom{m}{1} + \cdots + \binom{m}{k-1} + \frac{1+w}{2}\binom{m}{k}.$$

Now, by Proposition 4.5,

$$\mathrm{E}[\{H\} \cup \delta C] \geqslant 1 + \binom{m}{1} + \binom{m}{2} + \cdots + \binom{m}{k} + \frac{1+w}{2}\binom{m}{k+1},$$

where the union is disjoint since always $H \notin \delta C_j$. In terms of A and B, this implies

$$|A+B| \geqslant \frac{\binom{m+1}{0} + \binom{m+1}{1} + \cdots + \binom{m+1}{k} + \binom{m+1}{k+1} + w\binom{m}{k+1}}{2^{m+1}} \cdot |G|.$$

(3) What values can $m = \operatorname{codim} H$ take? By Lemma 2.5(vii), $\left(1 + \frac{m}{2}\right)/2^m \geqslant |A|/|G|$, where the case $m = 1$ is separately deduced from the assumption $|A|/|G| \leqslant 3/4$. On the other hand, by the theorem's assumption on t, $|A|/|G| > (t+2)/2^{t+1} = \left(1 + \frac{t}{2}\right)/2^t$. Since the sequence $\left(1 + \frac{n}{2}\right)/2^n$ is monotone, we infer $m < t$.

The theorem is obtained by plugging $m = t-1$ into the derived lower bound. We claim that for smaller m the bound is even higher, as demonstrated in Figure 3. Indeed, for each $1 \leqslant i \leqslant t-2$, the graph of the lower bound on $|A+B|$ given $m = i$ is concave down by the log-concavity of the binomial coefficients, $\binom{m}{k}/\binom{m}{k-1} \geqslant \binom{m}{k+1}/\binom{m}{k}$. Thus the graph of $m = i+1$, which connects the midpoints of adjacent segments in the $m = i$ graph, must be lower. □

https://doi.org/10.1017/S0963548312000351 Published online by Cambridge University Press

Remark. By the Hamming balls construction, the lower bound we have found is optimal on a biparametric discrete family of points. In view of our treatment of $F(K)$ in the previous section, we expect that at intermediate points better bounds should be provable.

There are three points where our approach to Theorem 1.1 may be suboptimal: the isoperimetric inequality we use is not always perfectly tight, the addition of $H \cup E$ instead of the whole of A, and dropping the assumption on B's affine span.

It is perhaps worth remarking that the machinery of compressions can still be applied under the assumption $\langle A \rangle = \langle B \rangle = G$. This is done by showing that without loss of generality we may assume that A and B are simultaneously compressed such that they include a common affine basis.

Here is a brief outline of how this is done. First, partition A and B into their intersections with cosets of $\langle (A-A) \cap (B-B) \rangle$. These parts can be translated without increasing $|A+B|$, such that $A-A$ and $B-B$ include a common basis of G. Then apply $\{i\}$-compressions with respect to this basis, until it is included in $A \cap B$.

Acknowledgement

I would like to thank my adviser, Professor Nati Linial, for his patient and helpful guidance during the research and preparation for this manuscript.

References

[1] Bollobás, B. and Leader, I. (1996) Sums in the grid. *Discrete Math.* **162** 31–48.
[2] Cauchy, A. L. (1813) Recherches sur les nombres. *J. École Polytechnique* **9** 99–123.
[3] Cohen, G. and Zémor, G. (1999) Subset sums and coding theory. *Astérisque* **258** 327–339.
[4] Davenport, H. (1935) On the addition of residue classes. *J. London Math. Soc.* **1** 30.
[5] Deshouillers, J. M., Hennecart, F. and Plagne, A. (2004) On small sumsets in $(\mathbb{Z}/2\mathbb{Z})^n$. *Combinatorica* **24** 53–68.
[6] Diao, H. (2009) Freiman–Ruzsa-type theory for small doubling constant. *Math. Proc. Camb. Phil. Soc.* **146** 269–276.
[7] Eliahou, S. and Kervaire, M. (1998) Sumsets in vector spaces over finite fields. *J. Number Theory* **71** 12–39.
[8] Eliahou, S. and Kervaire, M. (2005) Minimal sumsets in infinite abelian groups. *J. Algebra* **287** 449–457.
[9] Eliahou, S. and Kervaire, M. (2005) Old and new formulas for the Hopf–Stiefel and related functions. *Expositiones Mathematicae* **23** 127–145.
[10] Eliahou, S., Kervaire, M. and Plagne, A. (2003) Optimally small sumsets in finite abelian groups. *J. Number Theory* **101** 338–348.
[11] Frankl, P. (1984) A new short proof for the Kruskal–Katona theorem. *Discrete Math.* **48** 327–329.
[12] Frankl, P. (1987) The shifting technique in extremal set theory. In *Surveys in Combinatorics 1987* (C. Whitehead, ed.), Vol. 123 of *London Mathematical Society Lecture Notes*, Cambridge University Press, pp. 81–110.
[13] Frankl, P. (1989) A lower bound on the size of a complex generated by an antichain. *Discrete Math.* **76** 51–56.
[14] Freiman, G. A. (1973) *Foundations of a Structural Theory of Set Addition* (translated from the Russian), Vol. 37 of *Translations of Mathematical Monographs*, AMS.

[15] Gardner, R. J. and Gronchi, P. (2001) A Brunn–Minkowski inequality for the integer lattice. *Trans. Amer. Math. Soc.* **353** 3995–4024.
[16] Green, B. and Ruzsa, I. Z. (2006) Sets with small sumset and rectification. *Bull. London Math. Soc.* **38** 43.
[17] Green, B. and Tao, T. (2009) Freiman's theorem in finite fields via extremal set theory. *Combin. Probab. Comput.* **18** 335–355.
[18] Harper, L. H. (1966) Optimal numberings and isoperimetric problems on graphs. *J. Combin. Theory* **1** 385–393.
[19] Hennecart, F. and Plagne, A. (2003) On the subgroup generated by a small doubling binary set. *Europ. J. Combin.* **24** 5–14.
[20] Hopf, H. (1940) Ein toplogischer Beitrag zur reellen Algebra. *Comment. Math. Helv.* **13** 219–239.
[21] Katona, G. O. H. (1964) Intersection theorems for systems of finite sets. *Acta Math. Hungar.* **15** 329–337.
[22] Katona, G. O. H. (1975) The Hamming-sphere has minimum boundary. *Studia Sci. Math. Hungar.* **10** 131–140.
[23] Konyagin, S. V. (2008) On the Freiman theorem in finite fields. *Math. Notes* **84** 435–438.
[24] Kruskal, J. B. (1963) The number of simplices in a complex. In *Mathematical Optimization Techniques* (R. Bellman, ed.), University of California Press, pp. 251–278.
[25] Lev, V. F. (2003) Generating binary spaces. *J. Combin. Theory Ser. A* **102** 94–109.
[26] Lev, V. F. (2006) Critical pairs in abelian groups and Kemperman's structure theorem. *Internat. J. Number Theory* **2** 379–396.
[27] Pfjster, A. (1965) Zur Darstellung von -1 als Summe von Quadraten in einem Körper. *J. London Math. Soc.* **1** 159.
[28] Ruzsa, I. Z. (1992) Arithmetical progressions and the number of sums. *Periodica Math. Hungar.* **25** 105–111.
[29] Ruzsa, I. Z. (1994) Generalized arithmetical progressions and sumsets. *Acta Math. Hungar.* **65** 379–388.
[30] Ruzsa, I. Z. (1994) Sum of sets in several dimensions. *Combinatorica* **14** 485–490.
[31] Ruzsa, I. Z. (1999) An analog of Freiman's theorem in groups. *Astérisque* **258** 323–326.
[32] Sanders, T. (2008) A note on Freiman's theorem in vector spaces. *Combin. Probab. Comput.* **17** 297–305.
[33] Schneider, R. (1993) *Convex Bodies: The Brunn–Minkowski Theory*, Vol. 44, Cambridge University Press.
[34] Shapiro, D. B. (2000) *Compositions of Quadratic Forms*, De Gruyter.
[35] Stiefel, E. (1940) Über Richtungsfelder in den projektiven Räumen und einen Satz aus der reellen Algebra. *Comment. Math. Helv.* **13** 201–218.
[36] Tao, T. and Vu, V. (2006) *Additive Combinatorics*, Vol. 105 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press.
[37] Viola, E. (2007) Selected results in additive combinatorics: An exposition. In *Electronic Colloquium on Computational Complexity* (*ECCC*) **14**.
[38] Vosper, A. G. (1956) The critical pairs of subsets of a group of prime order. *J. London Math. Soc.* **1** 200.
[39] Yuzvinsky, S. (1981) Orthogonal pairings of Euclidean spaces. *Michigan Math. J.* **28** 131–145.
[40] Zémor, G. (1992) An extremal problem related to the covering radius of binary codes. In *Algebraic Coding*, Vol. 573 of *Lecture Notes in Computer Science*, Springer, pp. 42–51.
[41] Zémor, G. (1992) Subset sums in binary spaces. *Europ. J. Combin.* **13** 221–230.