Frankl-Rödl type theorems for codes and permutations

Peter Keevash * Eoin Long †

January 11, 2015

Abstract

We give a new proof of the Frankl-Rödl theorem on forbidden intersections, via the probabilistic method of dependent random choice. Our method extends to codes with forbidden distances, where over large alphabets our bound is significantly better than that obtained by Frankl and Rödl. We also apply our bound to a question of Ellis on sets of permutations with forbidden distances, and to establish a weak form of a conjecture of Alon, Shpilka and Umans on sunflowers.

1 Introduction

A family $\mathcal{A}$ of sets is said to be *l-avoiding* if $|A \cap B| \neq l$ for all $A, B \in \mathcal{A}$. Erdős conjectured (see [9]) that for any $\epsilon \in (0,1)$ there is $\delta = \delta(\epsilon) > 0$ such that given l with $\epsilon n \leq l \leq (1/2 - \epsilon)n$, any l-avoiding family $\mathcal{A} \subset \mathcal{P}[n]$ satisfies $|\mathcal{A}| \leq (2-\delta)^n$ and offered \$250 for a solution. In [14], Frankl and Rödl gave a positive answer to Erdős' conjecture, proving a stronger result for k-uniform l-avoiding families. Note that for $A, B \in \binom{[n]}{k}$, their intersection satisfies $\max(0, 2k-n) \leq |A \cap B| \leq k$. Now it is easily seen that if $l \geq k - o(n)$ or $l \leq \max(0, 2k-n) + o(n)$ then there exist l-avoiding families $\mathcal{A} \subset \binom{[n]}{k}$ with $|\mathcal{A}| = (1-o(1))^n \binom{n}{k}$. Frankl and Rödl showed that for l in between these extremes, all k-uniform l-avoiding families have exponentially small density.

*Mathematical Institute, Oxford OX2 6GG, UK. Email: Peter.Keevash@maths.ox.ac.uk.
Research supported in part by ERC grant 239696 and EPSRC grant EP/G056730/1.
†Mathematical Institute, Oxford OX2 6GG, UK. Email: Eoin.Long@maths.ox.ac.uk, eoinlong@post.tau.ac.il.
2010 Mathematics Subject Classification. Primary 05D05. Secondary 05D40, 94B65.

Theorem 1 (Frankl-Rödl)**.** *Let $\alpha, \epsilon \in (0,1)$ with $\epsilon \leq \alpha/2$. Let $k = \lfloor \alpha n \rfloor$ and $l \in [\max(0, 2k-n) + \epsilon n, k - \epsilon n]$. Then any l-avoiding family $\mathcal{A} \subset \binom{[n]}{k}$ satisfies $|\mathcal{A}| \leq (1-\delta)^n \binom{n}{k}$ where $\delta = \delta(\alpha, \epsilon) > 0$.*

Theorem 1 along with several extensions of the theorem proved in [14] have had a huge impact in a number of different areas including discrete geometry [15], communication complexity [20] and quantum computing [6].

In Section 2 of this paper we give a new proof of Theorem 1. We show that the theorem can in fact be deduced from an earlier theorem due to Frankl and Wilson (see Theorem 13 below). While our new proof of Theorem 1 does not seem to improve on the bounds given in [14], the same proof method does significantly improve bounds when we forbid distances over a larger underlying alphabet. Given $q \in \mathbb{N}$, $q \geq 2$, we will say that a set $\mathcal{C}$ is a q-ary code if $\mathcal{C} \subset [q]^n$. The Hamming distance between two words $x, y \in [q]^n$ is written as $d_H(x,y) = |\{i \in [n] : x_i \neq y_i\}|$. For a code $\mathcal{C}$ we write $d(\mathcal{C}) = \{d_H(x,y) : \text{ distinct } x, y \in \mathcal{C}\} \subset [n]$. Frankl and Rödl used Theorem 1 to prove the following result:

Theorem 2 (Frankl-Rödl)**.** *Let $\mathcal{C} \subset [q]^n$, and let ϵ satisfy $0 < \epsilon < 1/2$. Suppose that $\epsilon n < d < (1-\epsilon)n$, and d is even if $q = 2$. If $d \notin d(\mathcal{C})$, then $|\mathcal{C}| \leq (q-\delta)^n$ with some positive constant $\delta = \delta(\epsilon, q)$.*

(Note that, in order for Theorem 2 to hold for $q = 2$, we must have that d is even since the set $\mathcal{C}_0 = \{x \in \{0,1\}^n : \sum_i x_i \equiv 0 \pmod 2\}$ satisfies $|\mathcal{C}_0| = 2^{n-1}$ but contains no odd distances.)

In Section 3, we improve this to the following:

Theorem 3. *Let $\mathcal{C} \subset [q]^n$, and let ϵ satisfy $0 < \epsilon < 1/2$. Suppose that $\epsilon n < d < (1-\epsilon)n$, and d is even if $q = 2$. If $d \notin d(\mathcal{C})$, then $|\mathcal{C}| \leq q^{(1-\delta)n}$ with some positive constant $\delta = \delta(\epsilon)$.*

As a consequence of Theorem 3 we obtain a Frankl-Rödl type theorem for permutations. Given two permutations $\pi, \rho \in S_n$ we write

$$d_{S_n}(\pi, \rho) = |\{i \in [n] : \pi(i) \neq \rho(i)\}|.$$

For a set $\mathcal{S} \subset S_n$ we write

$$d_{S_n}(\mathcal{S}) = \{d \in [n] : d(\pi, \rho) = d \text{ for distinct } \pi, \rho \in \mathcal{S}\}.$$

Recently Ellis [10] asked how large a family $\mathcal{S} \subset S_n$ can be if $d \notin d_{S_n}(\mathcal{S})$ for some $d \in [n]$. A result of Deza and Frankl [8] answers this question for $d = n$,

showing that the largest such families have size $(n-1)!$. Ellis [10] gave a tight upper bound of $(n-2)!$ when $d = n-1$, provided n is sufficiently large. Here we consider this question when $\epsilon n < d < (1-\epsilon)n$ for $\epsilon > 0$. It is easily seen that for such d there exist sets of permutations $\mathcal{S} \subset S_n$ with $d \notin d_{S_n}(\mathcal{S})$ such that $|\mathcal{S}| \geq (n!)^c$ where $c = c(\epsilon) \in (0,1)$. By taking $q = n$ and viewing permutations $\pi \in S_n$ as vectors in $[q]^n$, with $\pi = (\pi(1), \ldots, \pi(n))$, since $|S_n| = n! = q^{(1-o(1))n}$, Theorem 3 has the following consequence:

Theorem 4. *Let $\mathcal{S} \subset S_n$, and let ϵ satisfy $0 < \epsilon < 1/2$. Suppose that $\epsilon n < d < (1-\epsilon)n$. If $d \notin d_{S_n}(\mathcal{S})$, then $|\mathcal{S}| < (n!)^{(1-\delta)}$ with some positive constant $\delta = \delta(\epsilon)$.*

Before we discuss another consequence of Theorem 3, we need the following definition.

Definition 5. *Given $v, w \in [q]^n$ let $Agree(v,w) = \{i \in [n] : (v)_i = (w)_i\}$. A collection of vectors $v_1, \ldots, v_k \in [q]^n$ is said to form a strong sunflower with k petals in $[q]^n$ if there is a fixed set $S \subset [n]$ such that $Agree(v_i, v_j) = S$ for all distinct $i, j \in [k]$. A collection of vectors $v_1, \ldots, v_k \in [q]^n$ is said to form a weak sunflower with k petals in $[q]^n$ if there is $D \in \mathbb{N}$ such that $|Agree(v_i, v_j)| = D$ for all distinct $i, j \in [k]$.*

Using Theorem 1, Frankl and Rödl proved that for any $k \in \mathbb{N}$ there exists $\delta = \delta(k) > 0$ such that if $\mathcal{A} \subset \{0,1\}^n$ with $|\mathcal{A}| > (2-\delta)^n$ then $\mathcal{A}$ contains a weak sunflower with k petals. Similarly, using the methods from [14] it can be shown that for any $k \in \mathbb{N}$ there exists $\delta = \delta(q,k) > 0$ so that given a code $\mathcal{C} \subset [q]^n$ with $|\mathcal{C}| \geq (q-\delta)^n$, $\mathcal{C}$ contains a weak k-petal sunflower in $[q]^n$. In Section 4 we prove the following:

Theorem 6. *Given $k \in \mathbb{N}$, there exists $\delta = \delta(k) > 0$ such that the following holds. For $q \geq 2$, every $\mathcal{C} \subset [q]^n$ which does not contain a weak sunflower with k petals satisfies $|\mathcal{C}| \leq q^{(1-\delta)n}$.*

This might be seen as giving evidence to a recent conjecture of Alon, Shpilka and Umans [2] who asked for a similar bound on families not containing a strong sunflower with 3 petals in $[q]^n$.

A crucial idea in the original proof of Theorem 1, along with an ingenious density increment argument, was to prove a stronger result. In [14] the authors actually proved a cross-intersecting version of Theorem 1:

Theorem 7 (Frankl-Rödl)**.** *Let $\alpha, \epsilon \in (0,1)$ with $\epsilon \leq \alpha/2$. Let $k = \lfloor \alpha n \rfloor$ and $l \in [\max(0, 2k-n) + \epsilon n, k - \epsilon n]$. Then if $\mathcal{A}_1, \mathcal{A}_2 \subset \binom{[n]}{k}$ with $|A_1 \cap A_2| \neq l$ for all $A_i \in \mathcal{A}_i$, they satisfy $|\mathcal{A}_1||\mathcal{A}_2| \leq (1-\delta)^n \binom{n}{k}^2$, where $\delta = \delta(\alpha, \epsilon) > 0$.*

We draw attention to the fact that the corresponding cross versions of Theorem 3 and Theorem 4 with our improved bounds do not hold in general. Indeed, for even n, if we take $\mathcal{A}_1$ to be the collection of all permutations in S_n sending $[n/2]$ to $[n/2]$ and $\mathcal{A}_2$ to be the collection of all permutations in S_n sending $[n/2]$ to $[n/2+1, n]$ we see that $|\mathcal{A}_i| \geq (n/2)!^2 \geq n!/3^n = (n!)^{1-o(1)}$ but $d_{S_n}(\rho_1, \rho_2) = n$ for all $\rho_i \in \mathcal{A}_i$.

However, in Section 5 we give a simple condition which guarantees fixed distances between such sets.

Theorem 8. *Given $\epsilon \in (0, 1/2)$ there exists $\delta', \gamma > 0$ such that the following holds. Let $q \geq 3$ and suppose that $\mathcal{C}, \mathcal{D} \subset [q]^n$ with $|\mathcal{C}| \geq q^{(1-\delta')n}$ and such that for all $x \in \mathcal{C}$ there exists $y \in \mathcal{D}$ with $d_H(x, y) \leq \gamma n$. Then given any $d \in (\epsilon n, (1-\epsilon)n)$, there exists $x \in \mathcal{C}$ and $y \in \mathcal{D}$ with $d_H(x, y) = d$.*

Using Theorem 8 in combination with some isoperimetric results, we recover a version of Frankl and Rödl's cross-distance result for fixed q.

Corollary 9. *Given $\epsilon \in (0, 1/2)$ and $q \geq 3$ there exists $\delta = \delta(\epsilon, q) > 0$ such that the following holds. Suppose that $\mathcal{C}, \mathcal{D} \subset [q]^n$ with $|\mathcal{C}||\mathcal{D}| > (q-\delta)^{2n}$. Then given any $d \in (\epsilon n, (1-\epsilon)n)$, there exists $x \in \mathcal{C}$ and $y \in \mathcal{D}$ with $d_H(x, y) = d$.*

Lastly, note that given $d \in [n]$ and any $x \in [q]^n$, there are exactly $\binom{n}{d}(q-1)^d$ words $y \in [q]^n$ with $d_H(x, y) = d$. In Section 6, we prove a supersaturated version of Theorem 3 (which is essentially best possible):

Theorem 10. *Given $\epsilon, \eta \in (0, 1/2)$ there is $\delta' > 0$ such that the following holds. Let $\mathcal{C} \subset [q]^n$ with $|\mathcal{C}| > q^{(1-\delta')n}$ and $d \in \mathbb{N}$ with $\epsilon n < d < (1-\epsilon)n$ (and d even if $q = 2$). Then there are at least $\binom{n}{d}(q-1)^d |\mathcal{C}| q^{-\eta n}$ pairs $x, y \in \mathcal{C}$ with $d_H(x, y) = d$.*

Notation: Given a set X, $\mathcal{P}(X)$ will denote the power set of X and $\binom{X}{k}$ will denote the collection of all subsets of size k in X. Given $m, n \in \mathbb{N}$ with $m \leq n$, $[n] = \{1, \ldots, n\}$ and $[m, n] = \{m, \ldots, n\}$. We also write $(n)_m$ for the falling factorial $(n)_m = n(n-1)\cdots(n-m+1)$.

2 Forbidding one intersection

In this section we give our new proof of Theorem 1. Before beginning, we outline the argument for the proof. In Lemma 12 below, we will show that given upper bounds on the size of k_i-uniform and l_i-avoiding families on n_i vertices for $i = 1, 2$, it is possible to obtain upper bounds on the size

of $(k_1 + k_2)$-uniform and $(l_1 + l_2)$-avoiding families on $(n_1 + n_2)$ vertices. The second tool we will need in the proof is the Frankl-Wilson theorem (see Theorem 13 and Corollary 14 below) which shows that the conclusion of Theorem 1 holds true for k-uniform, l-avoiding families whenever $k - l$ is a prime.

Suppose now that k, l and n are given as in the statement of Theorem 1, and assume that $k-l$ is odd. Let us choose $k_1, k_2, k_3 \in \mathbb{N}$ and $n_1, n_2, n_3 \in \mathbb{N}$ with $|k_i - k/3| < 1$ and $|n_i - n/3| < 1$ so that $k = \sum_{i=1}^{3} k_i$ and $n = \sum_{i=1}^{3} n_i$. We will also use a Vinogradov-type result due to Baker and Harman (see Theorem 15 below) which shows that any odd integer m can be written as a sum of three primes $m = a_1 + a_2 + a_3$, where $|a_i - m/3| = o(m)$. Applying this to $k - l$ we find three primes $k - l = a_1 + a_2 + a_3$ with $|a_i - (k-l)/3| = o(k-l) = o(n)$. Now set $l_i = k_i - a_i$ for all $i \in [3]$. Since $k_i - l_i = a_i$ is prime for all i, by the Frankl-Wilson theorem we know that Theorem 1 holds for k_i-uniform and l_i-avoiding families on n_i vertices, for all $i \in [3]$. Using Lemma 12 this will show that Theorem 1 holds for $(k_1 + k_2)$-uniform and $(l_1 + l_2)$-avoiding families on $(n_1 + n_2)$ vertices. By applying Lemma 12 once again, we obtain that Theorem 1 holds for k-uniform and l-avoiding families on n vertices, since $k = \sum_{i=1}^{3} k_i$, $n = \sum_{i=1}^{3} n_i$ and $l = \sum_{i=1}^{3} l_i$. The case when $k - l$ is even follows similarly, by writing $k - l$ as a sum of four primes.

To prove Lemma 12 we will make use of the probabilistic technique known as dependent random choice. The reader is directed to the recent survey of Fox and Sudakov [11] where many other interesting applications of the method are discussed. The following lemma gives a statement of the method which we will use in our applications. We include the short proof for convenience.

Lemma 11. *Suppose that $G = (X, Y, E)$ is a bipartite graph with $|X| = M, |Y| = N$ and $|E| = \alpha MN$. Then, for any $t \in \mathbb{N}$, there exists $X' \subset X$ with $|X'| \geq \alpha^t M/2$ with the property that for all $x_1, x_2 \in X'$ we have $|N_G(x_1) \cap N_G(x_2)| \geq \alpha M^{-1/t} N$.*

Proof. To begin choose uniformly at random t elements T with replacement from Y and let S denote the set of elements adjacent to all elements of T. By linearity of expectation

$$\mathbb{E}(|S|) = \sum_{x \in X} \Big(\frac{|N_G(x)|}{|Y|} \Big)^t \geq \alpha^t M$$

where the inequality follows from the convexity of the function $f(z) = z^t$.

We will say that a pair x, x' in S are *bad* if $|N_G(x) \cap N_G(x')| < \alpha M^{-1/t} N$. Now any bad pair has probability at most $(\frac{|N_G(x_1) \cap N_G(x_2)|}{N})^t \leq \alpha^t M^{-1}$ of appearing in S. Therefore, letting Z denote the number of bad pairs in S, we find that

$$\mathbb{E}(Z) \leq \alpha^t M^{-1} \binom{|X|}{2} \leq \alpha^t M/2.$$

In particular, $\mathbb{E}(|S| - Z) \geq \alpha^t M/2$. Fix a choice of T such that $|S| - Z$ is at least this big and delete one element from each bad pair x_1, x_2 in S. Taking X' to be the remaining set, we have $|X'| \geq |S| - Z \geq \alpha^t M/2$ and no pairs in X' are bad, as required. □

The next lemma shows how one can use Lemma 11 to build fixed intersections from smaller ones.

Lemma 12. *For $i = 1, 2$, suppose that $n_i, k_i, l_i \in \mathbb{N}$ and $p_i \in (0, 1)$ are such that any l_i-avoiding family $\mathcal{A}_i \subset \binom{[n_i]}{k_i}$ satisfies $|\mathcal{A}_i| \leq p_i \binom{n_i}{k_i}$. Suppose that $t \in \mathbb{N}$ satisfies $\binom{n_1}{k_1}^{-2} > p_2^t$. Then any $(l_1 + l_2)$-avoiding family $\mathcal{A} \subset \binom{[n_1+n_2]}{k_1+k_2}$ satisfies $|\mathcal{A}| \leq (2p_1)^{1/t} \binom{n_1+n_2}{k_1+k_2}$.*

Proof. Let $\mathcal{A} \subset \binom{[n_1+n_1]}{k_1+k_2}$ have size $|\mathcal{A}| = \alpha \binom{n_1+n_2}{k_1+k_2}$. We show that if $\alpha > (2p_1)^{1/t}$, then there exist $A, A' \in \mathcal{A}$ with $|A \cap A'| = l_1 + l_2$.

To begin, choose a partition of $[n_1 + n_2]$ uniformly at random into two sets V_1 and V_2 of size n_1 and n_2 respectively. Let $\mathcal{A}' \subseteq \mathcal{A}$ denote the set

$$\mathcal{A}' = \{A \in \mathcal{A} : |A \cap V_1| = k_1, |A \cap V_2| = k_2\}$$

and let Z denote the random variable $Z = |\mathcal{A}'|$. It is easy to see that $\mathbb{E}(Z) = \alpha \binom{n_1}{k_1} \binom{n_2}{k_2}$. We fix a partition $V_1 \cup V_2 = [n_1 + n_2]$ for which Z is at least this large.

Now we can view $\mathcal{A}'$ as the edge set of a bipartite graph $G = (X, Y, E)$ with vertex bipartition $X = \binom{V_1}{k_1}$ and $Y = \binom{V_2}{k_2}$ in which $AB \in E(G)$ when $A \cup B \in \mathcal{A}'$. We see that G has at least $\alpha |X||Y|$ edges. Apply Lemma 11 to G with t as in the statement to find a set $X' \subset X$ with $|X'| \geq \alpha^t |X|/2$ such that all distinct pairs $A_1, A_2 \in X'$ have at least $\alpha |X|^{-1/t} |Y|$ common neighbours in G. Now if $\alpha > (2p_1)^{1/t}$ then $|X'| > p_1 \binom{n_1}{k_1}$ and by definition of p_1, we find $A_1, A_2 \in X'$ with $|A_1 \cap A_2| = l_1$.

Let $\mathcal{B}'$ denote the set of common neighbours of A_1 and A_2 in G. By Lemma 11 we find that

$$|\mathcal{B}'| \geq \alpha |X|^{-1/t} |Y| > (2p_1)^{1/t} |X|^{-1/t} |Y| \geq \binom{n_1}{k_1}^{-2/t} |Y| > p_2 \binom{n_2}{k_2}.$$

The third inequality here holds since by definition of p_1 we have $p_1 \geq 1/\binom{n_1}{k_1}$ and the fourth holds as $\binom{n_1}{k_1}^{-2} > p_2^t$. But now by definition of p_2, there exists $B_1, B_2 \in \mathcal{B}'$ with $|B_1 \cap B_2| = l_2$. By construction it can be seen that we have $A_1 \cup B_1, A_2 \cup B_2 \in \mathcal{A}$ and clearly $|(A_1 \cup B_1) \cap (A_2 \cup B_2)| = l_1 + l_2$, as required. □

We will also make use of a theorem of Frankl and Wilson from [16].

Theorem 13 (Frankl-Wilson)**.** *Let $k, l \in \mathbb{N}$ such that $k-l$ is a prime power and $2l+1 \leq k$. Suppose that $\mathcal{A} \subseteq \binom{[n]}{k}$ is an l-avoiding family. Then $|\mathcal{A}| \leq \binom{n}{k-l-1}$.*

The following simple corollary of Theorem 13 will give us a slightly more convenient bound.

Corollary 14. *Let $\epsilon \in (0,1)$ and let $l, k \in \mathbb{N}$ with $l < k$, such that $k-l$ is prime with $\max(0, 2k-n) + \epsilon n < l < k - \epsilon n$. Then any l-avoiding family $\mathcal{A} \subset \binom{[n]}{k}$ satisfies $|\mathcal{A}| \leq c^n \binom{n}{k}$ where $c = c(\epsilon) < 1$.*

Proof. Let $\mathcal{A}$ be an l-avoiding family with $|\mathcal{A}| = \alpha\binom{n}{k}$. By averaging, there exists a set $T \in \binom{[n]}{l-\epsilon n}$ such that $\mathcal{A}_T = \{A \in \binom{[n]\setminus T}{k-|T|} : A \cup T \in \mathcal{A}\}$ has size $|\mathcal{A}_T| \geq \alpha\binom{n-|T|}{k-|T|}$. Setting $l' = \epsilon n$ and $k' = k - |T|$ it is easy to see that $\mathcal{A}_T$ is an l'-avoiding k'-uniform family. Since $k' = k-l+\epsilon n \geq 2\epsilon n + 1 = 2l'+1$ and $k'-l' = k-l$ is prime, by Theorem 13, we have $|\mathcal{A}_T| \leq \binom{n-|T|}{k'-l'} = \binom{n-|T|}{k-l}$. This gives that

$$\begin{aligned} \alpha &\leq \frac{\binom{n-|T|}{k-l}}{\binom{n-|T|}{k-|T|}} = \frac{(k-|T|)!(n-k)!}{(k-l)!(n-k+\epsilon n)!} \\ &= \frac{(k-l+\epsilon n)_{\epsilon n}}{(n-k+\epsilon n)_{\epsilon n}} \leq \Big(\frac{k-l+\epsilon n}{n-k+\epsilon n}\Big)^{\epsilon n} \leq \Big(\frac{1}{1+\epsilon}\Big)^{\epsilon n} \end{aligned}$$

since $n-k \geq k-l+\epsilon n$. Taking $c = (\frac{1}{1+\epsilon})^\epsilon < 1$, the result follows. □

Lastly, we will use the following Vinogradov-type result due to Baker and Harman [5] which says that every large enough odd number can be written as a sum of three primes of almost equal size.

Theorem 15 (Baker-Harman)**.** *Every odd integer $n > n_0$ can be written as a sum of three primes $n = a_1 + a_2 + a_3$ with $|a_i - n/3| \leq n^{4/7}$ for all i.*

Proof of Theorem 1. Let $\mathcal{A} \subset \binom{[n]}{k}$ be an l-avoiding family which satisfies $l \in [\max(0, 2k-n) + \epsilon n, k - \epsilon n]$. We wish to show that $|\mathcal{A}| \leq (1-\delta)^n \binom{n}{k}$, where $\delta = \delta(\alpha, \epsilon) > 0$. By taking δ to be sufficiently small, we may assume that the theorem holds for small values of $n \leq n_0 = n_0(\epsilon)$, so we will assume that $n \geq n_0$.

First suppose that $k-l$ is odd. Choose $k_1, k_2, k_3 \in \mathbb{N}$ and $n_1, n_2, n_3 \in \mathbb{N}$ with $\sum_{i=1}^3 k_i = k$ and $\sum_{i=1}^3 n_i = n$ with $|k_i - k/3| < 1$ and $|n_i - n/3| < 1$ for all i, with $n_1 \geq n_2 \geq n_3$. By Theorem 15, as $k - l > \epsilon n$ and $n > n_0(\epsilon)$, we can write $k - l = a_1 + a_2 + a_3$ where a_i is prime and $|(k-l)/3 - a_i| \leq \frac{\epsilon n}{8}$ for all i. Also set $l_i = k_i - a_i$ for all i. Then $k_i - l_i$ is prime for all i, $\sum_i k_i - l_i = k - l$ and $\max(0, 2k_i - n_i) + \epsilon n_i/2 \leq l_i \leq k_i - \epsilon n_i/2$.

By Corollary 14 any l_i-avoiding family $\mathcal{A}_i \subset \binom{[n_i]}{k_i}$ satisfies $|\mathcal{A}_i| \leq p_i \binom{n_i}{k_i}$ where $p_i = c_1^{n_i}$ with $c_1 = c(\epsilon/2) < 1$. Taking $t_1 = \lceil 2/\log_2(1/c_1) \rceil$ we find

$$p_2^{t_1} = c_1^{t_1 n_2} \leq 2^{-2n_1} < \binom{n_1}{k_1}^{-2}.$$

Therefore, by Lemma 12 any $(l_1 + l_2)$-avoiding family $\mathcal{B} \subset \binom{[n_1+n_2]}{k_1+k_2}$ with $|\mathcal{B}| = \beta \binom{n_1+n_2}{k_1+k_2}$ satisfies $\beta \leq (2c_1^{n_1})^{1/t_1}$.

To complete the proof we simply repeat the previous argument again. Let $t_2 = \lceil 4t_1/\log_2(1/c_1) \rceil$. Then we have

$$\beta^{t_2} \leq ((2c_1^{n_1})^{1/t_1})^{t_2} \leq (2c_1^{n_1})^{4/\log_2(1/c_1)} \leq 2^{-2n_3} < \binom{n_3}{k_3}^{-2}$$

where the third inequality holds since $n \geq n_0(\epsilon)$. Lemma 12 now gives that any l-avoiding family $\mathcal{A} \subset \binom{[n]}{k}$ satisfies $|\mathcal{A}| \leq c_2^n \binom{n}{k}$ where $c_2 = (c_1^{n_3})^{1/t_2 n} \leq c_1^{1/4t_2} < 1$. As c_1 and t_2 depend only on ϵ, this completes the proof in the case when $k-l$ is odd.

The case where $k-l$ is even can be proved by splitting $k-l$ into 4 primes of almost equal size. The proof now proceeds identically to the odd case, using an additional application of Lemma 12. □

3 Forbidding code distances

In this section we prove Theorem 3. We will assume that $q \geq 3$ throughout the section, as the case $q = 2$ follows from Theorem 1.

Given a set $V \subset [n]$, a vector $x \in [q]^V$ will be indexed by elements of V, i.e. $x = (x_i)_{i \in V}$. Given two disjoint sets V and W and vectors $x \in [q]^V$ and

$y \in [q]^W$, we write $x \circ y$ for the vector $x \circ y \in [q]^{V \cup W}$, the *concatenation* of x and y, given coordinatewise by

$$(x \circ y)_i = \begin{cases} x_i & \text{if } i \in V \\ y_i & \text{if } i \in W. \end{cases}$$

Given a collection of disjoint sets $V_1, \ldots, V_k$ and vectors $x_i \in [q]^{V_i}$ for all $i \in [k]$, we will write $x_1 \circ x_2 \cdots \circ x_k \in [q]^{\cup_i V_i}$ to be the vector

$$x_1 \circ x_2 \cdots \circ x_k = (((x_1 \circ x_2) \circ x_3) \circ \cdots \circ x_k).$$

(Note, this notation does not depend on the order that the x_i are taken in).

We also require the following definition:

Definition 16. *Given a prime p and a set $\mathcal{D} \subset \mathbb{Z}_p \setminus \{0\}$, we say that a code $\mathcal{C} \subset [q]^n$ is a $\mathcal{D}$ (mod p)-code if for all $d \in d(\mathcal{C})$, we have $d \equiv d' \mod p$ for some $d' \in \mathcal{D}$.*

The following theorem, due to Frankl [12] (see also [4]), gives an upper bound on the size of (mod p)-codes .

Theorem 17 (Frankl)**.** *Suppose that p is a prime and that $\mathcal{C} \subset [q]^n$ is a $\mathcal{D}$ (mod p)-code with $|\mathcal{D}| = l$. Then $|\mathcal{C}| \leq \sum_{i=0}^{l} \binom{n}{i}(q-1)^i$.*

In applying Theorem 17 we use the following estimate due to Chernoff [7]. Let $q \in \mathbb{N}$ with $q \geq 3$. Then given $\alpha \in (0, (q-1)/q)$, we have

$$S_q(\alpha, n) := \sum_{i=0}^{\alpha n} \binom{n}{i}(q-1)^i \leq q^{f_q(\alpha)n}$$

where $f_q(\alpha) = \alpha \log_q(\frac{q-1}{\alpha}) + (1-\alpha)\log_q(\frac{1}{1-\alpha})$.

Proposition 18. *For $q \geq 3$ and $\alpha \in [0, 3/5]$ we have $S_q(\alpha, n) \leq q^{(1-1/125)n}$.*

Proof. First note the following:

(i) $\frac{\partial f_q}{\partial \alpha}(\alpha) = \log_q\left[\frac{(q-1)(1-\alpha)}{\alpha}\right] \geq 0$ for $\alpha \in [0, (q-1)/q]$;

(ii) $\frac{\partial^2 f_q}{\partial \alpha^2}(\alpha) = \frac{1}{\log_e q}\left[-\frac{1}{1-\alpha} - \frac{1}{\alpha}\right] \leq 0$, so $f_q(\alpha)$ is concave as a function of α on $[0,1]$. As $f_q(0) = 0$ and $f_q(\frac{q-1}{q}) = 1$, this shows that $f_q(\alpha) \geq \frac{q\alpha}{q-1}$ for $\alpha \in [0, (q-1)/q]$;

(iii) $\frac{\partial f_q}{\partial q}(\alpha) = \frac{1}{q \log_e q}\left[\frac{q\alpha}{q-1} - f_q(\alpha)\right] \leq 0$ for $\alpha \in [0, (q-1)/q]$ by (ii).

But then, for $q \geq 3$ and $\alpha \in [0, 3/5] \subset [0, (q-1)/q]$, we have

$$f_q(\alpha) \leq f_3(\alpha) \leq f_3(3/5) \leq 0.992,$$

where the first inequality holds since $f_q(\alpha)$ is decreasing in q by (iii), the second since $f_3(\alpha)$ is increasing in α by (i) and the third by a numerical calculation. □

Combined with Proposition 18, Theorem 17 now gives the following corollary.

Corollary 19. *Let $\epsilon \in (0,1)$ and $q \geq 3$. Suppose that p is a prime with $\epsilon n < p < 3n/5$ and that $\mathcal{C} \subset [q]^n$ is a code with $p \notin d(\mathcal{C})$. Then $|\mathcal{C}| \leq q^{(1-\delta_1)n}$ where $\delta_1 = \delta_1(\epsilon) > 0$.*

Proof. Suppose that $|\mathcal{C}| = \alpha q^n$. Choose t so that $p \in (\frac{n-t}{2}, \frac{3(n-t)}{5})$ – this is possible by the stated bound on p above. Now given a set $T \in \binom{[n]}{t}$ and elements $a_i \in [q]$ for $i \in T$, let

$$\mathcal{C}_T = \{x \in \mathcal{C} : x_i = a_i \text{ for all } i \in T\}.$$

By averaging we find $T \in \binom{[n]}{t}$ and $\{a_i \in [q] : i \in T\}$ such that $|\mathcal{C}_T| \geq \alpha q^{n-t}$. View $\mathcal{C}_T$ as a subset of $[q]^{[n]\setminus T}$. Clearly $p \notin d(\mathcal{C}_T)$. Since $p > (n-t)/2$, the set $\mathcal{C}_T$ is a $\mathcal{D} \pmod p$ code in $[q]^{[n]\setminus T}$, where $\mathcal{D} = \{1, \ldots, p-1\}$. Therefore by Theorem 17 and Proposition 18

$$\alpha q^{n-t} \leq |\mathcal{C}_T| \leq S_q(3/5, n-t) \leq q^{(1-1/125)(n-t)}.$$

Therefore $\alpha \leq q^{-(n-t)/125} \leq q^{-\epsilon n/125}$ using $\epsilon n \leq p \leq n-t$. Taking $\delta_1(\epsilon) = \epsilon/125$ completes the proof. □

Corollary 19 will allow us to deal with forbidden distances which are not too large. For larger distances we will use the following diametric theorem for $[q]^n$ due to Ahlswede and Khachatrian [1]. The diameter of a set $\mathcal{C} \subset [q]^n$, $\operatorname{diam}(\mathcal{C})$, is defined as

$$\operatorname{diam}(\mathcal{C}) := \max\{d : d \in d(\mathcal{C})\}.$$

Given $t \in \mathbb{N}$ and $r \in \mathbb{N} \cup \{0\}$, let $\mathcal{K}_r(t) \subset [q]^n$ denote the set

$$\mathcal{K}_r(t) = \{v \in [q]^n : |\{i \in [t+2r] : v_i = 1\}| \geq t + r\}.$$

It is easy to see that $\operatorname{diam}(\mathcal{K}_r(t)) = n - t$ for all r. The following remarkable result shows that given q and t, for some r, $\mathcal{K}_r(t)$ is the largest code in $[q]^n$ with diameter $n - t$.

Theorem 20 (Ahlswede, Khachatrian)**.** *Let $q, t \in \mathbb{N}$ with $q \geq 2$ and let $r \in \mathbb{N} \cup \{0\}$ be the largest integer such that*

$$t + 2r < \min\Big\{n + 1, t + 2\frac{t-1}{q-2}\Big\}. \tag{1}$$

Then any code $\mathcal{C} \subset [q]^n$ with $\operatorname{diam}(\mathcal{C}) \leq n - t$ satisfies $|\mathcal{C}| \leq |\mathcal{K}_r(t)|$. (By convention, $(t-1)/(q-2) = \infty$ if $q = 2$.)

We will use the following simple consequence of Theorem 20.

Corollary 21. *Given $\epsilon \in (0, 1/3)$ and $q \in \mathbb{N}$ with $q \geq 3$, every set $\mathcal{C} \subset [q]^n$ with $\operatorname{diam}(\mathcal{C}) \leq (1-\epsilon)n$ satisfies $|\mathcal{C}| \leq q^{(1-\delta_2)n}$ where $\delta_2 = \delta_2(\epsilon) > 0$.*

Proof. Let $t = \epsilon n$. Since $\epsilon < 1/3$, we have

$$n + 1 > \epsilon n + 2\epsilon n - 1 \geq t + 2\frac{t-1}{q-2},$$

so the minimum in (1) is attained by the right hand term and gives $r = \lceil (t-1)/(q-2) \rceil - 1$ in Theorem 20. Therefore to prove the statement, by Theorem 20 it suffices to prove that $|\mathcal{K}_r(t)| \leq q^{(1-\delta_2)n}$. We have

$$\begin{aligned} |\mathcal{K}_r(t)| &= \Big(\sum_{i=0}^{r}(q-1)^i\binom{t+2r}{i}\Big)q^{n-t-2r} = S_q\big(\frac{r}{t+2r}, t+2r\big)q^{n-t-2r} \\ &\leq q^{(1-1/125)(t+2r)}q^{n-t-2r} = q^{n-(t+2r)/125} \leq q^{(1-\epsilon/125)n}, \end{aligned}$$

using Proposition 18 in the first inequality and that $\epsilon n \leq t < t + 2r$ in the second. Taking $\delta_2(\epsilon) = \epsilon/125$ completes the proof. □

The next lemma is an analogue of Lemma 12 for subsets of $[q]^n$ and can be proved similarly.

Lemma 22. *For $i = 1, 2$, suppose that $n_i, d_i \in \mathbb{N}$ and $p_i \in (0, 1)$ are such that if $\mathcal{C}_i \subset [q]^{n_i}$ with $d_i \notin d(\mathcal{C}_i)$ then $|\mathcal{C}_i| \leq p_i q^{n_i}$. Suppose that $t \in \mathbb{N}$ satisfies $q^{-2n_1} > p_2^t$. Then any set $\mathcal{C} \subset [q]^{n_1+n_2}$ with $d_1 + d_2 \notin d(\mathcal{C})$ satisfies $|\mathcal{C}| \leq (2p_1)^{1/t}q^{n_1+n_2}$.*

We are now ready for the proof of Theorem 3.

Proof of Theorem 3. Let $\mathcal{C} \subset [q]^n$ with $|\mathcal{C}| = \alpha q^n$ where $q \geq 3$ and suppose that for some $d \in [\epsilon n, (1-\epsilon)n]$ we have $d \notin d(\mathcal{C})$. We wish to show that $\alpha \leq q^{-\delta n}$ where $\delta = \delta(\epsilon) > 0$. By taking δ sufficiently small, we can assume

that the result holds for $n < n_0(\epsilon)$, so we will assume that $n \geq n_0$. The proof will split into two cases, according as $d \in [\epsilon n, \frac{11}{20}n]$ or $d \in [\frac{11}{20}n, (1-\epsilon)n]$.

Case 1: $d \in [\epsilon n, \frac{11}{20}n]$

We will suppose that d is odd, as the case of even d is similar. As $d \geq \epsilon n$, provided $n \geq n_0(\epsilon)$, Theorem 15 allows us to write $d = d_1 + d_2 + d_3$ where d_i are primes with $|d_i - d/3| \leq \epsilon n/100$. Also, partition n as a sum of naturals $n = n_1 + n_2 + n_3$ where $|n_i - n/3| \leq 1$ with $n_1 \geq n_2 \geq n_3$. For $n \geq n_0(\epsilon)$ this gives that for all $i \in [3]$ we have

$$d_i \geq d/3 - \epsilon n/100 \geq \epsilon n/3 - \epsilon n/100 \geq \epsilon n_i/2 \tag{2}$$

and

$$d_i \leq d/3 + \epsilon n/100 \leq 11n/60 + \epsilon n/100 < 3n_i/5. \tag{3}$$

Now set $V_1 = [n_1]$, $V_2 = [n_1 + 1, n_1 + n_2]$ and $V_3 = [n_1 + n_2 + 1, n]$. For each $i = 1, 2, 3$ we have that d_i is prime and by (2) and (3) it satisfies $\epsilon|V_i|/2 \leq d_i \leq 3|V_i|/5$. This gives that the hypotheses of Corollary 19 are satisfied (taking $\epsilon/2$ in place of ϵ) and therefore any code $\mathcal{C}_i \subset [q]^{V_i}$ with $d_i \notin d(\mathcal{C}_i)$ satisfies $|\mathcal{C}_i| \leq p_i q^{n_i}$ where $p_i = q^{-\delta_1(\epsilon/2)n_i}$.

We now use these bounds to obtain an upper bound on the size of codes $\mathcal{B} \subset [q]^{V_1 \cup V_2}$ with $d_+ := d_1 + d_2 \notin d_H(\mathcal{B})$ using Lemma 22. We first claim that taking $t_1 = \lceil 4/\delta_1(\epsilon/2) \rceil$, we have $q^{-2n_1} > p_2^{t_1}$. Indeed,

$$p_2^{t_1} = q^{-\delta_1(\epsilon/2)n_2 t_1} \leq q^{-\delta_1(\epsilon/2)n_2(\frac{4}{\delta_1(\epsilon/2)})} = q^{-4n_2} < q^{-2n_1},$$

since $n_2 > n_1/2$. Lemma 22 then shows that any code $\mathcal{B} \subset [q]^{V_1 \cup V_2}$ with $d_+ \notin d(\mathcal{B})$ satisfies $|\mathcal{B}| \leq \alpha_1 q^{n_1+n_2}$ where $\alpha_1 = (2q^{-\delta_1(\epsilon/2)n_1})^{1/t_1}$.

To complete the proof we now repeat the previous argument to obtain a bound for codes $\mathcal{C} \subset [q]^{V_1 \cup V_2 \cup V_3} = [q]^n$ with $d = d_+ + d_3 \notin d_H(\mathcal{C})$. We first claim that setting $t_2 = \lceil 4t_1/\delta_1(\epsilon/2) \rceil$ we have $\alpha_1^{t_2} < q^{-2n_3}$. Indeed,

$$\alpha_1^{t_2} = (2q^{-\delta_1(\epsilon/2)n_1})^{t_2/t_1} \leq (2q^{-\delta_1(\epsilon/2)n_1})^{4/\delta_1(\epsilon/2)} = 2^{4/\delta_1(\epsilon/2)} q^{-4n_1} < q^{-2n_3}.$$

The last inequality holds since $2^{4/\delta_1(\epsilon/2)} < 2^{2n/3} \leq q^{2n/3} \leq q^{2n_1}$ for $n \geq n_0(\epsilon) \geq 2/\delta_1(\epsilon/2)$.

Now any code $\mathcal{B} \subset [q]^{V_1 \cup V_2}$ with $d_+ \notin d_H(\mathcal{B})$ satisfies $|\mathcal{B}| \leq \alpha_1 q^{n_1+n_2}$ and any code $\mathcal{C}_3 \subset [q]^{V_3}$ with $d_3 \notin d(\mathcal{C}_3)$ satisfies $|\mathcal{C}_3| \leq p_3 q^{n_3}$ where $p_3 = q^{-\delta_1(\epsilon/2)n_3}$ and $\alpha_1^{t_2} < q^{-2n_3}$. Therefore by Lemma 22 any code $\mathcal{C} \subset [q]^{V_1 \cup V_2 \cup V_3} = [q]^n$ with $d = d_+ + d_3 \notin d(\mathcal{C})$ satisfies $|\mathcal{C}| \leq \alpha_2 q^n$ where

$$\begin{aligned}\alpha_2 &= (2p_3)^{1/t_2} \leq (2q^{-\delta_1(\epsilon/2)n_3})^{1/t_2} = 2^{1/t_2} q^{-\delta_1(\epsilon/2)n_3/t_2} \\ &\leq 2q^{-\delta_1(\epsilon/2)n/4t_2} \leq q^{-\delta_1(\epsilon/2)n/8t_2} = q^{-\delta_3(\epsilon)n}\end{aligned}$$

where $\delta_3(\epsilon) := \delta_1(\epsilon/2)/8t_2$. The last inequality here holds since for $n > n_0(\epsilon)$ we have $2 \leq q^{\delta_1(\epsilon/2)n/8t_2}$.

Case 2: $d \in [\frac{11}{20}n, (1-\epsilon)n]$

We will prove this case using Case 1 above. The proof will work as follows. We will choose a partition $[n] = V_1 \cup V_2$ and use dependent random choice to find a large subset $X \subset [q]^{V_1}$ such that for each pair of elements $x, x' \in X$, the set of *common extensions* $y \in [q]^{V_2}$ with $x \circ y, x' \circ y \in \mathcal{C}$ is also large. We can then apply Corollary 21 to find a pair $x, x' \in X$ at Hamming distance $d' \approx |V_1|$. Provided $|V_1|$ is carefully chosen, this will ensure that $d - d' \leq 3|V_2|/5$. We may apply then apply Case 1 to the set of extensions of x and x' to find a pair y, y' with $d_H(y, y') = d - d'$. Then $x \circ y, x' \circ y' \in \mathcal{C}$ lie at Hamming distance d.

Let δ_3 denote the same function as in Case 1 and let δ_2 denote the function in Corollary 21. Choose $n_1 \in [n]$ such that

$$|\frac{29}{40}n_1 + \frac{11}{40}n - d| \leq 1. \tag{4}$$

As $d \geq \frac{11}{20}n$, this gives $n/4 \leq n_1 \leq (1-\epsilon)n$. Take $t = \lceil 2/\epsilon\delta_3(1/4)\rceil$ and $\delta_4(\epsilon) = \delta_2(\epsilon/4)/8t > 0$. We will show that if $\mathcal{C} \subset [q]^n$ where $|\mathcal{C}| = \alpha q^n$ with $\alpha > q^{-\delta_4(\epsilon)n}$ then $\mathcal{C}$ contains two words at Hamming distance d.

To begin, partition $[n] = V_1 \cup V_2$ where $V_1 = [n_1]$ and $V_2 = [n_1 + 1, n]$. We set $n_2 = n - n_1 = |V_2|$. As in Lemma 12, view the elements of $\mathcal{C}$ as edges of a bipartite graph $G = (X, Y, E)$ with bipartition $X = [q]^{V_1}$ and $Y = [q]^{V_2}$, where $xy \in E(G)$ if $x \circ y \in \mathcal{C}$. Clearly $|E(G)| = \alpha|X||Y|$. Apply Lemma 11 to G with t as above to find a set $X' \subset X$ with

$$\begin{aligned} |X'| = \alpha^t|X|/2 &> q^{-\delta_4(\epsilon)tn}q^{n_1}/2 = q^{-\delta_2(\epsilon/4)n/8}q^{n_1}/2 \\ &\geq q^{-\delta_2(\epsilon/4)n/4}q^{n_1} \geq q^{(1-\delta_2(\epsilon/4))n_1} \end{aligned}$$

(using $q^{\delta_2(\epsilon/4)n/8} \geq 2$ for $n \geq n_0(\epsilon)$ and $n_1 \geq n/4$) such that all distinct x, x' in X' share at least $\alpha|X|^{-1/t}|Y|$ common neighbours in Y. By Corollary 21 (with $\epsilon/4$ in place of ϵ) there exists $x, x' \in X'$ with $d_H(x, x') = d'$, with d' satisfying $(1 - \epsilon/4)n_1 \leq d' \leq n_1$. Let $\mathcal{B} \subset [q]^{V_2}$ denote the set of common extensions of x, x' in Y. We have

$$\begin{aligned} \log_q|\mathcal{B}| &\geq \log_q(\alpha|X|^{-1/t}|Y|) > \log_q(q^{-\delta_4(\epsilon)n}q^{-n_1/t}q^{n_2}) \\ &= -\delta_4(\epsilon)n - n_1/t + n_2 > -2n_1/t + n_2 \\ &\geq -2n_1\epsilon\delta_3(1/4)/2 + n_2 = -\epsilon\delta_3(1/4)n_1 + n_2 \\ &= -\delta_3(1/4)n_2\Big(\frac{\epsilon n_1}{n_2}\Big) + n_2 \geq (1 - \delta_3(1/4))n_2. \end{aligned}$$

The final inequality used $n_1/n_2 = n_1/(n-n_1) \leq 1/\epsilon$ since $n_1 \leq (1-\epsilon)n$. Therefore $|\mathcal{B}| > q^{(1-\delta_3(1/4))n_2}$. As $d' \in [(1-\epsilon/4)n_1, n_1]$ and from (4) we have $|(d-n_1) - \frac{11}{40}n_2| = |\frac{29}{40}n_1 + \frac{11}{40}n - d| \leq 1$, we find

$$\begin{aligned} d - d' &\in [d-n_1, d-(1-\epsilon/4)n_1] \\ &= [d-n_1, (d-n_1)+\epsilon n_1] \\ &\subset [\frac{11}{40}n_2 - 1, \frac{11}{40}n_2 + \epsilon n_1/4] \\ &\subset [\frac{1}{4}n_2, \frac{11}{20}n_2]. \end{aligned}$$

Here we again used $n_1/n_2 = n_1/(n-n_1) \leq 1/\epsilon$. This shows that $\mathcal{B} \subset [q]^{V_2}$ satisfies $|\mathcal{B}| > q^{(1-\delta_3(1/4))n_2}$. By definition of δ_3, there exists a pair $y, y' \in \mathcal{B}$ with $d_H(y,y') = d-d'$. But this gives $x \circ y, x' \circ y' \in \mathcal{C}$. As $d_H(x \circ y, x' \circ y') = d$ this completes the proof of this case.

Taking $\delta(\epsilon) = \min(\delta_3(\epsilon), \delta_4(\epsilon))$ completes the proof of the Theorem. $\square$

4 Weak sunflowers in $[q]^n$

In this section, we will prove Theorem 6. For convenience, we will assume that n is a multiple of k with $n = km$; this assumption can easily be removed. Set $V_i = [(i-1)m+1, im]$ for all $i \in [k]$. We will prove by induction on k that given $\epsilon > 0$ and $d \in [\epsilon m, (1-\epsilon)m]$ (with d even if $q = 2$), there exists $\delta' = \delta'(\epsilon, k) > 0$ with the following property: for any set $\mathcal{C} \subset [q]^n$ with $|\mathcal{C}| > q^{(1-\delta')n}$, there exists $x_i, y_i \in [q]^{V_i}$ for $i \in [k]$ with $d_H(x_i, y_i) = d$, such that $z_1 \circ \cdots \circ z_k \in \mathcal{C}$ for any choice of $z_i \in \{x_i, y_i\}$. This will complete the proof as taking

$$v_i = x_1 \circ \cdots \circ x_{i-1} \circ y_i \circ x_{i+1} \circ \cdots \circ x_k,$$

the set $\{v_1, \ldots, v_k\}$ is a weak-sunflower with k petals contained in $\mathcal{C}$.

The case when $k = 1$ follows immediately from Theorem 3, so we will assume by induction that the result holds for $k-1$ and prove it for k. Let $W_1 = \cup_{i=2}^k V_i$ so that $[n] = V_1 \cup W_1$. Letting $t = \lceil 2/((k-1)\delta'(\epsilon, k-1)) \rceil$, we claim that we can take $\delta' = \delta'(\epsilon, k) = \delta(\epsilon)/2kt$, where $\delta(\epsilon)$ is as in Theorem 3. Indeed, as in the proof of Lemma 12, view elements of $[q]^n$ as edges of a bipartite graph $G = ([q]^{V_1}, [q]^{W_1}, E)$ in which $xy \in E(G)$ if $x \circ y \in \mathcal{C}$. Then if $|\mathcal{C}| = |E(G)| = \alpha q^n$ where $\alpha > q^{-\delta' n}$, by Lemma 11, there exists a set $\mathcal{C}_1 \subset [q]^{V_1}$ with

$$|\mathcal{C}_1| \geq \alpha^t q^m/2 \geq q^{-\delta(\epsilon)n/2k} q^m/2 = q^{(1-\delta(\epsilon)/2)m}/2 \geq q^{(1-\delta(\epsilon))m}$$

with any two elements in $\mathcal{C}_1$ sharing at least

$$\alpha q^{-m/t} q^{(k-1)m} > q^{-\frac{\delta(\epsilon)n}{2kt} - \frac{m}{t}} q^{(k-1)m} \geq q^{-\frac{2m}{t}} q^{(k-1)m} \geq q^{(1-\delta(\epsilon,k-1))(k-1)m}$$

common neighbours in G. But then, by Theorem 3, $\mathcal{C}_1$ must contain elements x_1 and x_2 with $d_H(x_1, x_2) = d$. Also, by the induction hypothesis for $k-1$, we find $x_i, y_i \in [q]^{V_i}$ for all $i \in [2,k]$ with $d_H(x_i, y_i) = d$ such that all elements of the set

$$\{z_2 \circ \cdots \circ z_k : z_i \in \{x_i, y_i\} \text{ for all } i \in [2,k]\}$$

are common neighbours of both x_1 and y_1. But by definition of G, this means that $z_1 \circ \cdots \circ z_l \in \mathcal{C}$ for any choice of $z_i \in \{x_i, y_i\}$ for all $i \in [k]$, as claimed.

5 Forbidding distances between pairs of sets in $[q]^n$

Proof of Theorem 8. Given ϵ we will take $\delta'(\epsilon) = \delta(\epsilon/2)/2$, where $\delta(\epsilon/2)$ is as in Theorem 3 and $\gamma = \min(\epsilon/2, \frac{\delta(\epsilon/2)}{16\log(1/\delta(\epsilon/2))})$. Let $q \geq 3$ and suppose that $\mathcal{C}, \mathcal{D} \subset [q]^n$ with $|\mathcal{C}| \geq q^{(1-\delta')n}$ and such that for all $x \in \mathcal{C}$ there is $y \in \mathcal{D}$ with $d_H(x,y) \leq \gamma n$. Suppose $d \in (\epsilon n, (1-\epsilon)n)$. We will show that there exists $x \in \mathcal{C}$ and $y \in \mathcal{D}$ with $d_H(x,y) = d$.

From the statement, for all $x \in \mathcal{C}$ there is some $y_x \in \mathcal{D}$ with $d_H(x, y_x) \leq \gamma n$. By pigeonholing, there must be a set $T \subset \binom{[n]}{\gamma n}$ and a subset $\mathcal{C}' \subset \mathcal{C}$ with $|\mathcal{C}'| \geq |\mathcal{C}|/\binom{n}{\gamma n} \geq |\mathcal{C}|2^{-H(\gamma)n}$ with the property that, for all $x \in \mathcal{C}'$, we have $\{i \in [n] : (x)_i \neq (y_x)_i\} \subset T$. There are at most $q^{\gamma n}$ choices for both $x|_T$ and $y_x|_T$, so again by pigeonholing we find $\mathcal{C}'' \subset \mathcal{C}'$ with $|\mathcal{C}''| \geq |\mathcal{C}'|/q^{2\gamma n}$ and vectors $f_0, g_0 \in [q]^T$ such that $x|_T = f_0$ and $y_x|_T = g_0$ for all $x \in \mathcal{C}''$. Let $d_H(f_0, g_0) = t \leq \gamma n \leq \epsilon n/2$. Now by choice of γ, we have $H(\gamma) \leq \delta(\epsilon/2)/4$ and $\gamma < \delta(\epsilon/2)/8$ and so

$$|\mathcal{C}''| \geq |\mathcal{C}|2^{-H(\gamma)n} q^{-2\gamma n} \geq q^{(1-\delta(\epsilon/2))n}.$$

Therefore, since $\epsilon n/2 \leq d - \gamma n \leq d - t \leq (1-\epsilon)n$ by Theorem 3 there are $x, x' \in \mathcal{C}''$ with $d_H(x, x') = d - t$. But then

$$d_H(x', y_x) = \underbrace{d_H(x, y_x)}_{\text{distance in } T} + \underbrace{d_H(x', x)}_{\text{distance in } [n] \setminus T} = d_H(f_0, g_0) + d_H(x', x) = d.$$

As $x' \in \mathcal{C}$ and $y_x \in \mathcal{D}$, this completes the proof. □

We will now show that the conditions of Theorem 8 can be applied to $\mathcal{C}, \mathcal{D} \subset [q]^n$ with $|\mathcal{C}|, |\mathcal{D}| \geq (q-\delta)^n$ provided $\delta = \delta(\epsilon, q) > 0$ is sufficiently small, proving Corollary 9. Let K_q^n denote the graph on vertex set $[q]^n$, in which $x, y \in [q]^n$ are adjacent if they differ on a single coordinate. Given a set $\mathcal{C} \subset [q]^n$, let $N^{(t)}(\mathcal{C}) = \{x \in [q]^n : d_H(x, x') \leq t \text{ for some } x' \in \mathcal{C}\}$. The following result gives an approximate vertex isoperimetric theorem for K_q^n (see [18] or [3]).

Theorem 23. *For all $q \geq 2$ and $\gamma > 0$, there exists $\delta'' = \delta''(\gamma, q) > 0$ such that the following holds. Any set $\mathcal{C} \subset [q]^n$ with $|\mathcal{C}| \geq (q-\delta'')^n$ satisfies $|N^{(\gamma n)}(\mathcal{C})| > q^n - (q-\delta'')^n$.*

Proof of Corollary 9. Let $q \geq 3$ and $\epsilon > 0$. We wish to show that, given any $\mathcal{C}, \mathcal{D} \subset [q]^n$ with $|\mathcal{C}||\mathcal{D}| > (q-\delta)^{2n}$ where $\delta = \delta(\epsilon, q) > 0$ is sufficiently small, for any $d \in (\epsilon n, (1-\epsilon)n)$ there exists $x \in \mathcal{C}$ and $y \in \mathcal{D}$ with $d_H(x, y) = d$. By taking δ to be sufficiently small it suffices to prove the result for $n \geq n_0(\epsilon, q)$. Given our value of ϵ, let $\delta' > 0$ and $\gamma > 0$ be as in Theorem 8. We will take $\delta = \delta''(\gamma, q)/4$, with $\delta''(\gamma, q)$ as in Theorem 23.

First note that as $|\mathcal{C}|, |\mathcal{D}| \leq q^n$ we have $|\mathcal{C}|, |\mathcal{D}| \geq \frac{(q-\delta)^{2n}}{q^n} > (q-2\delta)^n$. Take $\mathcal{C}' \subset \mathcal{C}$ be the set $\mathcal{C}' = \{x \in \mathcal{C} : \exists y \in \mathcal{D} \text{ with } d_H(x, y) \leq \gamma n\}$. By Theorem 8, it suffices to show that $|\mathcal{C}'| > q^{(1-\delta')n}$. Suppose for contradiction that this is not the case. Then for $n \geq n_0$, we have

$$|\mathcal{C} \setminus \mathcal{C}'| \geq |\mathcal{C}| - |\mathcal{C}'| > (q-2\delta)^n - q^{(1-\delta')n} \geq (q-4\delta)^n = (q-\delta'')^n.$$

Now by Theorem 23 we have $|N^{(\gamma n)}(\mathcal{C} \setminus \mathcal{C}')| > q^n - (q-\delta'')^n$. However, by definition of $\mathcal{C}'$ we have $\mathcal{D} \cap N^{(\gamma n)}(\mathcal{C} \setminus \mathcal{C}') = \emptyset$. As $|\mathcal{D}| \geq (q-2\delta)^n = (q-\delta''/2)^n > (q-\delta'')^n$ this is a contradiction and it completes the proof of the corollary. □

6 Supersaturated version of Theorem 3

In this section we prove Theorem 10 which gives a supersaturated version of Theorem 3. A rough outline of the proof is as follows. We will assume for contradiction that we are given a large set $\mathcal{C}$, as in Theorem 10, which contains few pairs at Hamming distance d, for some $\epsilon n \leq d \leq (1-\epsilon)n$. By fixing some coordinate entries from $[n]$ and restricting others, we will find a large subset $\mathcal{E}' \subset [r]^{V_1}$ with $r \leq q$ and $(1+\alpha')d \leq |V_1| \leq n$ which contains no pairs at distance d. However, if $\mathcal{E}'$ is a sufficiently large subset $[r]^{V_1}$, this will contradict Theorem 3.

A technical complication which occurs in our proof is that on reducing from $[q]^n$ to $[r]^{V_1}$, it is possible that $q \geq 3$ and d is odd, but that after restricting we end up with $r = 2$. Note that in this case, we cannot apply Theorem 3 to $\mathcal{E}'$, as only even distances can be guaranteed in $[2]^{V_1}$. To get around this obstacle, we first prove the case when d is even separately, and then deduce the odd case from it by an application of dependent random choice.

Proof of Theorem 10. To begin, we will set $\alpha = \eta/(16\log(16/\eta))$ and $\delta' = \eta\epsilon\delta(\alpha/2)/8$, where δ is as in Theorem 3. Also set $m = \alpha n$ and $r = \max\{\lfloor q^{\eta/4} \rfloor, 2\}$. Let $\mathcal{C} \subset [q]^n$ with $|\mathcal{C}| > q^{(1-\delta')n}$. We will show that given d with $\epsilon n \leq d \leq (1-\epsilon)n$, the code $\mathcal{C}$ contains at least $\binom{n}{d}(q-1)^d|\mathcal{C}|q^{-\eta n}$ pairs $x, y \in \mathcal{C}$ with $d_H(x,y) = d$. As discussed above, we start by giving the proof in the case where d is even.

Let N denote the number of pairs $\{x, y\}$ with $x, y \in \mathcal{C}$ such that $d_H(x,y) = d$. Make the following selection of random choices:

- choose a partition of $[n] = V_1 \cup V_2$ with $|V_1| = d+m$ and $|V_2| = n-d-m$ uniformly at random;
- for each $i \in V_1$, choose a subset $Q_i \subset [q]$ of size r uniformly at random;
- for each $i \in V_2$, choose an element $q_i \in [q]$ uniformly at random.

We will say that an element $x \in \mathcal{C}$ is a *captured element* if $x_i \in Q_i$ for all $i \in V_1$ and $x_j = q_j$ for all $j \in V_2$. Let $\mathcal{E} \subset \mathcal{C}$ denote the set of captured elements. We also say that a pair $\{x, y\} \in \mathcal{C}^{(2)}$ is a *captured d-pair* if $x, y \in \mathcal{E}$, $d_H(x,y) = d$ and $V_2 \subset \mathrm{Agree}(x,y)$.

Let X and Y denote the random variables which count the number of captured elements and the number of captured d-pairs respectively. Clearly, given $x \in \mathcal{C}$, we have $\mathbb{P}(x \in \mathcal{E}) = r^{d+m}/q^n$. Therefore we have

$$\mathbb{E}(X) = \frac{r^{d+m}|\mathcal{C}|}{q^n}. \tag{5}$$

For a fixed pair $x, y \in \mathcal{C}$ with $d_H(x,y) = d$ we have

$$\mathbb{P}(x, y \text{ form a captured } d\text{-pair}) = \frac{\binom{d+m}{d}}{\binom{n}{d}} \left(\frac{\binom{r}{2}}{\binom{q}{2}} \right)^d \left(\frac{r}{q} \right)^m q^{-(n-d-m)}$$

To see this we justify one term at a time. The first term is the probability that the d coordinates on which x and y differ are included in V_1. The second

term is the probability that the entries x_i and y_i are included in Q_i where they differ. The third term is the probability that the remaining (common) entries of $x_i = y_i$ in V_1 are included in Q_i and the last term is the probability that $q_i = x_i$ for $i \in V_2$.

Suppose for contradiction that $N < \binom{n}{d}(q-1)^d|\mathcal{C}|q^{-\eta n}$. Then

$$\begin{aligned}\mathbb{E}(Y) &= \Big(\frac{r}{q}\Big)^m\Big(\frac{r(r-1)}{q(q-1)}\Big)^d q^{-(n-d-m)}\frac{\binom{d+m}{m}}{\binom{n}{d}}N \\ &< \binom{d+m}{m}\frac{r^{d+m}(r-1)^d}{q^n}|\mathcal{C}|q^{-\eta n} \\ &\leq \Big(\frac{\binom{n}{m}(r-1)^n}{q^{\eta n}}\Big)\frac{r^{d+m}|\mathcal{C}|}{q^n}.\end{aligned}$$

Now for $\alpha \leq \eta/(16\log(16/\eta))$ we have $H(\alpha) \leq \eta/4$ and so $\binom{n}{m} \leq 2^{H(\alpha)n} < q^{\eta n/4}$. Since we also have $(r-1) \leq \max\{q^{\eta/4}, 1\} = q^{\eta/4}$ we have

$$\mathbb{E}(Y) \leq \frac{1}{q^{\eta n/2}}\frac{r^{d+m}|\mathcal{C}|}{q^n} \leq \frac{r^{d+m}|\mathcal{C}|}{2q^n}$$

for $n \geq n_0 \geq 2/\eta$. Combined with (5), as $|\mathcal{C}| \geq q^{(1-\delta')n}$, this gives

$$\mathbb{E}(X-Y) \geq \frac{r^{d+m}|\mathcal{C}|}{2q^n} \geq \frac{r^{d+m}q^{-\delta' n}}{2}. \tag{6}$$

But $q^{-\delta' n} = (q^{\eta/4})^{-\delta(\alpha/2)\epsilon n/2} \geq r^{-\delta(\alpha/2)\epsilon n/2} \geq r^{-\delta(\alpha/2)(d+m)/2}$ as $d \geq \epsilon n$. This shows that

$$\mathbb{E}(X-Y) \geq \frac{r^{(1-\delta(\alpha/2)/2)(d+m)}}{2} > r^{(1-\delta(\alpha/2))(d+m)}. \tag{7}$$

The second inequality here holds since $r^{\delta(\alpha/2)(d+m)/2} \geq r^{\delta(\alpha/2)\epsilon n/2} \geq 2$ for $n \geq n_0$. Fix choices of V_1, V_2, Q_i for all $i \in V_1$ and q_i for $i \in V_2$ such that $X - Y$ is at least this big. Now remove one element from every captured d-pair in $\mathcal{E}$. By (7), this leaves a set $\mathcal{E}' \subset \mathcal{E}$ with

$$|\mathcal{E}'| > r^{(1-\delta(\alpha/2))(d+m)} \tag{8}$$

which contains no captured d-pairs.

Now $\mathcal{E}'$ is a subset of $\prod_{i\in V_1} Q_i \times \prod_{j\in V_2}\{q_j\}$ and this product set is naturally identified with $[r]^{d+m}$. We also have

$$\frac{\alpha}{2}(d+m) \leq d \leq \big(1-\frac{\alpha}{2}\big)(d+m).$$

Indeed, $\alpha(d+m)/2 \leq d$ since $m \leq d$ and $\alpha \leq 1$ and $d \leq (1-\alpha/2)(d+m)$ since $\alpha d/2 \leq \alpha n/2 = m/2 \leq (1-\alpha/2)m$. But now since $\mathcal{E}'$ does not contain a pair (x,y) with $d_H(x,y) = d$, by Theorem 3 we have $|\mathcal{E}'| \leq q^{(1-\delta(\alpha/2))(d+m)}$. However this contradicts (8). Therefore we must have $N \geq \binom{n}{d}(q-1)^d|\mathcal{C}|q^{-\eta n}$, as required. This completes the proof of the case of d even.

The case of d odd and $q \geq 3$ can be deduced from the even case by dependent random choice. As it is similar to previous steps, we only outline the argument. Given ϵ and η as in the statement we will assume that γ and ζ are chosen sufficiently small so that various estimates against functions of ϵ and η hold. First choose a partition of $[n] = V_1 \cup V_2$, where $|V_1| = \gamma n$. By dependent random choice (see Lemma 11), provided that $|\mathcal{C}| \geq q^{(1-\delta)n}$ with $\delta = \delta(\gamma,\zeta) > 0$ sufficiently small, there is a subset $\mathcal{S} \subset [q]^{V_1}$ with $|\mathcal{S}| \geq q^{(1-\zeta)|V_1|}$ such that for every pair $x, y \in \mathcal{S}$, the set

$$\mathcal{C}_{x,y} := \{z \in [q]^{V_2} : \text{ with } x \circ z, y \circ z \in \mathcal{C}\}$$

satisfies $|\mathcal{C}_{x,y}| \geq q^{(1-\zeta)|V_2|}$. Now as ζ is small, we can apply Theorem 3 to $\mathcal{S}$ to find $x, y \in \mathcal{S}$ with $d_H(x,y) = d'$ odd — this is possible since $q \geq 3$. Taking $d'' = d - d'$, as $d' \leq |V_1| = \gamma n$, we have

$$\epsilon|V_2|/2 < \epsilon n - \gamma n \leq d'' \leq (1-\epsilon)n \leq (1-\epsilon/2)|V_2|,$$

provided γ was chosen with $\gamma < \epsilon/4$.

We now use the even case proven above to guarantee that $\mathcal{C}_{x,y}$ contains many pairs at Hamming distance d''. Indeed, as $|\mathcal{C}_{x,y}| \geq q^{(1-\zeta)|V_2|}$ and d'' is even, if ζ was chosen so that $\zeta < \delta'(\epsilon/2, \eta/2)$, the even case shows that $\mathcal{C}_{x,y}$ contains at least $\binom{|V_2|}{d''}(q-1)^{d''}|\mathcal{C}_{x,y}|q^{-\eta|V_2|/2}$ pairs at Hamming distance d''. Expanding we find

$$\begin{aligned}\binom{|V_2|}{d''}(q-1)^{d''}|\mathcal{C}_{x,y}|q^{-\eta|V_2|/2} &\geq \binom{n}{d}\frac{(d)_{d'}}{(n)_{d'}} \times (q-1)^d q^{-d'} \\ &\quad \times |\mathcal{C}|q^{-\zeta n - \gamma n} \times q^{-\eta n/2} \\ &\geq \binom{n}{d}(q-1)^d|\mathcal{C}|\big(\frac{\epsilon}{2}\big)^{\gamma n}q^{-(\zeta+2\gamma+\eta/2)n}.\end{aligned}$$

The second inequality here holds since $\frac{(d)_{d'}}{(n)_{d'}}q^{-d'} \geq (\epsilon/2)^{\gamma n}q^{-\gamma n}$ holds for $d' \leq \gamma n < \epsilon n/2$. For γ and ζ sufficiently small in comparison with ϵ and η, this gives at least $\binom{n}{d}(q-1)^d|\mathcal{C}|q^{-\eta n}$ such pairs $z, z' \in \mathcal{C}_{x,y}$. But for each of these pairs, we have $x \circ z, y \circ z' \in \mathcal{C}$ and $d_H(x \circ z, y \circ z') = d' + d'' = d$. This completes the proof. □

7 Concluding Remarks

In this paper we gave improved bounds on the size of codes and families of permutations with a forbidden distance. These bounds demonstrate the power of dependent random choice in forbidden distance problems and we expect that the method will have many more applications in extremal set theory.

It remains an intriguing open problem to obtain a better upper bound on the size of maximum l-avoiding families $\mathcal{A} \subset \mathcal{P}[n]$. A natural construction is to take all sets that are 'large' or 'small', where 'large' sets have size at least $(n+l)/2$ and 'small' sets have size less than l. (If $n+l$ is odd we can also add all sets of size $(n+l-1)/2$ containing 1). For fixed l and large n, Frankl and Füredi [13] proved that this is the unique extremal family.

However, much less is known when l is comparable with n. Under the stronger condition of being $(l+1)$-intersecting, Katona [17] showed that the family of all large sets gives the optimal construction. Mubayi and Rödl [19] conjectured that for the l-avoiding problem, with any $\epsilon n < l < (1/2 - \epsilon)n$, the same family of all large sets and all small sets as before should be approximately optimal, say up to a multiplicative factor of $2^{o(n)}$. They proved this when the l-avoiding condition is replaced with the stronger condition of having a small forbidden interval of intersections around l.

Acknowledgements

We would like to thank both referees for their careful reading of the paper and for several helpful comments which improved the presentation.

References

[1] R. Ahlswede, L.H. Khachatrian: The diametric theorem in Hamming spaces - optimal anticodes, *Adv. Appl. Math.* **20**(4) (1998), 429-449.

[2] N. Alon, A. Shpilka, C. Umans: On sunflowers and matrix multiplication, *Comput. Complexity* **22**(2) (2013), 219-243.

[3] B. Bollobás, I. Leader: Compressions and isoperimetric inequalities, *J. Combin. Theory Ser. A* **56**(1) (1991), 47-62.

[4] L. Babai, H. Snevily, R.M. Wilson: A new proof of several inequalities on codes and sets, *J. Combin. Theory Ser. A* **71**(1) (1995), 146-153.

[5] R.C. Baker, G. Harman: The three primes theorem with almost equal primes, *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **356** (1998), no. 1738, 763-780.

[6] H. Buhrman, R. Cleve, A. Wigderson: Quantum vs. classical communication and computation, *Proceedings of 30th STOC* (1998), 63-68.

[7] H. Chernoff: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statistics* **23**(4) (1952), 493-507.

[8] M. Deza, P. Frankl: On the maximum number of permutations with given maximal or minimal distance, *J. Combin. Theory Ser. A* **22**(3) (1977), 352-360.

[9] P. Erdős: Problems and results in graph theory and combinatorial analysis, *Proc. Fifth British Comb. Conf. 1975 Aberdeen*, Congressus Numerantium, 15-Utilitas Math., Winnipeg, 1976.

[10] D. Ellis: Forbidding just one intersection, for permutations, *J. Combin. Theory Ser. A* **126** (2014), 136-165.

[11] J. Fox, B. Sudakov: Dependent random choice, *Random Structures Algorithms* **38**(1-2) (2011), 68-99.

[12] P. Frankl: Orthogonal vectors in the n-dimensional cube and codes with missing distances, *Combinatorica* **6**(3) (1986), 279-285.

[13] P. Frankl, Z. Füredi: On hypergraphs without two edges meeting in a given number of vertices, *J. Combin. Theory Ser. A* **36**(2) (1984), 230-236.

[14] P. Frankl, V. Rödl: Forbidden intersections, *Trans. Amer. Math. Soc.* **300**(1) (1987), 259-286.

[15] P. Frankl, V. Rödl: A partition property of simplices in euclidean space, *J. Amer. Math. Soc.* **3**(1) (1990), 1-7.

[16] P. Frankl, R.M. Wilson: Intersection theorems with geometric consequences, *Combinatorica* **1**(4) (1981), 357-368.

[17] G.O.H. Katona: Intersection theorems for systems of finite sets, *Acta Math. Acad. Sci. Hung.* **15** (1964), 329-337.

[18] L.H. Harper: On an isoperimetric problem for Hamming graphs, *Discrete Appl. Math.* **95**(1-3) (1999), 285-309.

[19] D. Mubayi, V. Rödl: Specified intersections, *Trans. Amer. Math. Soc.* **366**(1) (2014), 491-504.

[20] J. Sgall, Bounds on pairs of families with restricted intersections, *Combinatorica* **19**(4), (1999), 555-566.