Forbidden intersections for codes

Keevash, Peter; Lifshitz, Noam; Long, Eoin; Minzer, Dor

DOI:
10.1112/jlms.12801

Received: 4 November 2021 | Revised: 19 June 2023 | Accepted: 21 June 2023

DOI: 10.1112/jlms.12801

RESEARCH ARTICLE

Journal of the London Mathematical Society

Forbidden intersections for codes

Peter Keevash[1] | **Noam Lifshitz**[2] | **Eoin Long**[3] | **Dor Minzer**[4]

[1]Mathematical Institute, University of Oxford, UK

[2]Einstein Institute of Mathematics, Hebrew University, Jerusalem, Israel

[3]School of Mathematics, University of Birmingham, Birmingham, UK

[4]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA

Correspondence
Dor Minzer, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA.
Email: minzer.dor@gmail.com

Funding information
ERC, Grant/Award Number: 883810

Abstract

Determining the maximum size of a t-intersecting code in $[m]^n$ was a longstanding open problem of Frankl and Füredi, solved independently by Ahlswede and Khachatrian and by Frankl and Tokushige. We extend their result to the setting of forbidden intersections, by showing that for any $m > 2$ and n large compared with t (but not necessarily m) that the same bound holds for codes with the weaker property of being $(t-1)$-avoiding, that is, having no two vectors that agree on exactly $t-1$ coordinates. Our proof proceeds via a junta approximation result of independent interest, which we prove via a development of our recent theory of global hypercontractivity: we show that any $(t-1)$-avoiding code is approximately contained in a t-intersecting junta (a code where membership is determined by a constant number of coordinates). In particular, when $t = 1$, this gives an alternative proof of a recent result of Eberhard, Kahn, Narayanan and Spirkl that symmetric intersecting codes in $[m]^n$ have size $o(m^n)$.

MSC 2020
05D05, 05D10, 06E30 (primary)

1 | INTRODUCTION

Many intersection problems for finite sets (see the survey [13]) have natural generalisations to a setting variously described as codes, vectors or integer sequences. For example, any intersecting family of subsets of $[n]$ has size at most 2^{n-1}, and more generally, any intersecting code in $[m]^n$ has size at most m^{n-1}, where we say that a code $\mathcal{F} \subset [m]^n$ is intersecting if for any x, y in $\mathcal{F}$, there is some i with $x_i = y_i$. However, these settings are quite different, in that there are many maximum intersecting families of sets, including very symmetric examples such as the family of all sets of size $> n/2$, whereas in $[m]^n$ for $m > 2$, the only example is obtained by fixing one coordinate to have a fixed value. A more substantial difference was recently demonstrated by Eberhard, Kahn, Narayanan and Spirkl [5], who showed that adding a symmetry assumption reduces the maximum size to $o(m^n)$.

A longstanding open problem of Frankl and Füredi [9] posed the corresponding question for codes $\mathcal{F} \subset [m]^n$ that are *t-intersecting*, in that any x, y in $\mathcal{F}$ have *agreement*$\mathrm{agr}(x, y) = |\{i : x_i = y_i\}| \geqslant t$. From the perspective of coding theory, one may think of such $\mathcal{F}$ as an 'anti-code', in that we are imposing an upper bound on the Hamming distance between any two of its vectors. From a combinatorial perspective, the natural analogy is with t-intersecting k-graphs (k-uniform hypergraphs), for which the extremal question was also a longstanding open problem, posed by Erdős, Ko and Rado [8] and finally resolved by the Complete Intersection Theorem of Ahlswede and Khachatrian [1]. The analogous result for codes, resolving the problem of Frankl and Füredi, was also obtained by Ahlswede and Khachatrian [2], and independently by Frankl and Tokushige [12]. They showed that the maximum size of a t-intersecting code in $[m]^n$ is achieved by one of the following natural examples, which can be thought of as Hamming balls on a subset of the coordinates, and which we will simply call 'balls' (following [28]): let

$$\mathcal{S}_{t,r}[m]^n = \{x \in [m]^n : |\{j \in [1, t+2r] : x_j = 1\}| \geqslant t + r\}.$$

We show for any $m > 2$ and n large compared with t (but not necessarily m) that the same conclusion holds under the weaker assumption that $\mathcal{F}$ is $(t-1)$-*avoiding*, that is, no x, y in $\mathcal{F}$ have agreement $t - 1$.

Theorem 1.1. *For all $t \in \mathbb{N}$, there is $n_0 \in \mathbb{N}$ such that if $\mathcal{F} \subset [m]^n$ is a $(t-1)$-avoiding code with $m \geqslant 3$ and $n \geqslant n_0$, then $|\mathcal{F}| \leqslant \max_{r \geqslant 0} |\mathcal{S}_{t,r}[m]^n|$ with equality only when $\mathcal{F}$ is isomorphic to a ball.*

Theorem 1.1 can be viewed as an analogue for codes of the classical forbidden intersection problem for set systems, which has a substantial literature, particularly stemming from the many applications of the celebrated Frankl–Rödl theorem [11] (see also [10, 17]). Our proof (discussed in the next subsection) proceeds via a junta approximation result of independent interest, showing that any $(t-1)$-avoiding code is approximately contained in a t-intersecting junta (a code where membership is determined by a constant number of coordinates). In particular, when $t = 1$, this gives an alternative proof of the result of [5], as a family that essentially depends on few coordinates is very far from being symmetric.

1.1 | Overview of the proof

The proof of Theorem 1.1 has three steps, each of which has elements of independent interest.

(1) Junta approximation: any $(t-1)$-avoiding code is approximately contained in a t-intersecting junta.
(2) Anticode stability: a stability version of the Ahlswede–Khachatrian theorem on anticodes determines the structure of the junta from (1) — it must be a certain ball $\mathcal{F}$.

(3) Bootstrapping: given that the code of maximum size is close to $\mathcal{F}$, it must in fact be equal to $\mathcal{F}$.

The methods required to implement these three steps depend considerably on the size of m, and we need a variety of ideas in Combinatorics and Analysis, some of which are new. The most significant new idea in this paper is a random gluing operation, which may be thought of as a natural, more versatile, analogue of the sharp threshold phenomenon from the biased hypercube, as we explain below.

We remark that techniques that go into the proof of Theorem 1.1 and the various steps above are quite flexible and are useful to study the structure of extremal families with respect to more general classes of forbidden configurations. In particular, in Section 8, we establish junta approximation results for a richer class of forbidden configurations (rather than just the configuration consisting of two points that agree on $t-1$ coordinates). Such results are often strong enough to get approximate versions of Theorem 1.1; however, nailing down the exact size of an extremal family is a more challenging task that we leave for future research.

Random gluings

Often times, when working over the p-biased Boolean hypercube, that is, $\{0,1\}^n$ along with the measure $\mu_p(x) = p^{|\{i\in[n]\ |x_i=1\}|}(1-p)^{|\{i\in[n]\ |x_i=0\}|}$, one is interested in studying the structure of a monotone family $\mathcal{F} \subseteq \{0,1\}^n$ (i.e. a family such that if $x \in \mathcal{F}$ and $x_i \leqslant y_i$ for all i, then $y \in \mathcal{F}$). One particularly useful idea is to see how much the measure of the family changes when increasing p, that is, study the behaviour of $\mu_p(\mathcal{F}) = \Pr_{x\sim\mu_p}[x \in \mathcal{F}]$ as a function of p. It is easy to see that this is an increasing function of p, and the main point of this idea is that the rate of increase tells us a lot about the structure $\mathcal{F}$ has. In a nutshell, unless the family $\mathcal{F}$ has some local, junta-like, structure, this increase must be sharp.[†] This idea plays significant role is various problems in analysis and extremal combinatorics, but seems to be specific to the cube: one heavily relies on an ordering of $\{0,1\}^n$ which makes sense with respect to intersection problems, and such orderings do not exist on many other domains, such as $[m]^n$.

Our random gluing operator may be viewed as a natural extension of the above operator to $[m]^n$, which is also potentially more versatile and may be relevant in other domains. Given a $k < m$ and a family $\mathcal{F} \subseteq [m]^n$, we think of shrinking the alphabet (in each coordinate independently) from m to k, by identifying each symbol $\sigma \in [m]$ with a symbol from $[k]$, that is, given such identifications $\pi_i : [m] \to [k]$ for each i, one may consider the family $\mathcal{F}_\pi = \{(\pi_1(x_1), \dots, \pi_n(x_n))\ |x \in \mathcal{F}\}$. It is clear that such operation is 'friendly' with respect to intersection problems (e.g. if $\mathcal{F}$ is t-intersecting, then so is $\mathcal{F}_\pi$). We show that this operation, when sampling $\pi_1, \dots, \pi_n$ appropriately and considering an appropriate product measure on $[k]^n$, also enjoys the second effect of the 'increasing p' idea from above. Namely, we show that unless $\mathcal{F}$ has local structure (i.e. if $\mathcal{F}$ is global as per Definition 5.2), one can find a gluing operation that increases the measure of $\mathcal{F}$ significantly.

The analysis of this gluing operation proceeds via noise stability and a new hypercontractive inequality in general product spaces, which further extends our recent theory of global hypercontractivity introduced in [16]. This part of the argument can also be viewed as a development of the Junta Method (see [4, 16, 19].)

[†] When p is bounded away from 0 and 1, this structure is simply a junta, but when $p = o(1)$ or $p = 1 - o(1)$, this structure may be more complicated and is not fully understood (some partial results are known [14, 16, 22]). The notion of 'local structure' in this case considered herein corresponds to having restrictions of the family $\mathcal{F}$ with significant measure.

The following is a precise statement of our junta approximation theorem, which is a stability theorem of independent interest, describing the approximate structure of any $(t-1)$-avoiding code with size that is within a constant factor of the maximum possible.

Theorem 1.2. *For every $t \in \mathbb{N}$ and $\eta > 0$, there are n_0 and J in $\mathbb{N}$ such that if $\mathcal{F} \subset [m]^n$ is a $(t-1)$-avoiding code with $m \geqslant 3$ and $n \geqslant n_0$, then there is a t-intersecting J-junta $\mathcal{J} \subset [m]^n$ such that $|\mathcal{F} \setminus \mathcal{J}| \leqslant \eta|\mathcal{J}|$.*

As mentioned above, Theorem 1.2 implies the result of [5], as a junta is far from being symmetric. The assumption $m \geqslant 3$ is necessary, as when $m = 2$, we have symmetric examples as mentioned above. When $m > m_0(t)$ is large, we will in fact obtain a more precise statement: $\mathcal{J}$ will be a subcube of co-dimension t and we will give effective estimates for the approximation parameter η (see Theorems 6.3 and 7.14).

Our first ingredient in the proof of Theorem 1.2 is a regularity lemma, showing that any code can be approximately decomposed into a constant number of pieces, each of which is pseudorandom, in a certain sense that depends on the size of m. When $m < m_0(t)$ is fixed and $n > n_0(t, m)$ is large, each piece is such that constant size restrictions cannot significantly affect the measure. This is a strong pseudorandomness condition, from which the proof can be completed fairly easily using a result of Mossel on Markov chains hitting pseudorandom sets [24]. The idea is that if two restrictions defining the regularity decomposition agree in fewer than t coordinates, then we can impose a further restriction to make them agree in exactly $t-1$ coordinates, with no significant loss in measure by pseudorandomness. If our code is $(t-1)$-avoiding these restrictions must be cross intersecting, but Mossel's result implies that this is impossible for pseudorandom codes of non-negligible measure.

When m is large, one cannot obtain such a strong pseudorandomness condition in a regularity lemma, so we settle for the weaker property of uncapturability. A family $\mathcal{F} \subseteq [m]^n$ is said to be uncapturable if it is not approximately contained in a union of constantly many 'dictatorships', that is, families of the form $D_{i\to j} = \{x \in [m]^n \mid x_i = j\}$ for $i \in [n]$ and $j \in [m]$. We stress here that m is not thought of as constant, so one cannot fix i and take $D_{i\to j}$ for all $j \in [m]$. Our regularity lemma in this case shows that any given family $\mathcal{F}$ may be decomposed into pieces, such that each piece is uncapturable. This weaker regularity lemma makes it significantly harder to establish the t-intersection property as outlined above in the case that m is fixed; the main issue is that uncapturability may not be preserved by further restrictions.

Furthermore, if m is 'huge' (by which we will mean exponential in n), then the cross-agreement statement used for fixed m is false. To see this, consider the codes $\mathcal{E}$ having all vectors with all coordinates even, and $\mathcal{O}$ having all vectors with all coordinates odd. There is no non-zero agreement between $\mathcal{E}$ and $\mathcal{O}$, yet they are both highly uncapturable, and have measure 2^{-n} (which is non-negligible when m is huge).

The above example naturally suggests a further case: we say that m is 'moderate' if it is large but not huge. In this case, the high-level proof strategy is the same as for fixed m, although the required cross-agreement statement for uncapturable codes is difficult to prove, and this is where we need the most significant new ideas of the paper (gluing and global hypercontractivity). On the other hand, when m is huge, the above example shows that we need a different proof strategy. Here we draw inspiration from more combinatorial arguments of Keller and Lifshitz [19] which we adapt to the setting of codes by thinking of $\mathcal{F} \subset [m]^n$ as an n-partite n-graph (n-uniform hypergraph) with parts of size m. While the high-level strategy is similar to that in [19], the implementation is quite different; for example, the key to bootstrapping in this case turns out to be a subtle application of Shearer's entropy inequality.

We write $\mathcal{S}_{n,m,t}$ for a largest family among $\{\mathcal{S}_{t,r}[m]^n : r \geqslant 0\}$. From Theorem 1.2, we see that if a $(t-1)$-avoiding code $\mathcal{F} \subset [m]^n$ is at least as large as $\mathcal{S}_{n,m,t}$, then it is close to a t-intersecting junta. This raises the stability question for t-intersecting codes, which is the second ingredient in our proof of Theorem 1.1: must this junta be close to an extremal family? When m is large compared with t, it is not hard to show that such a junta must be close to a subcube of co-dimension t, that is, the ball $\mathcal{S}_{t,0}[m]^n$. For fixed m, the picture is more complex, and the full range of balls can occur; nevertheless, we are able to establish the required stability version of the Ahlswede–Khachatrian anticode theorem.

Theorem 1.3. *For every $t \in \mathbb{N}$ and $\varepsilon > 0$, there is $\delta > 0$ such that if $\mathcal{F} \subset [m]^n$ is t-intersecting with $m \geqslant 3$ and $|\mathcal{F}| \geqslant (1-\delta)|\mathcal{S}_{n,m,t}|$, then $|\mathcal{F} \setminus \mathcal{S}| \leqslant \varepsilon|\mathcal{S}|$ for some family $\mathcal{S}$ which is isomorphic to $\mathcal{S}_{n,m,t} = \mathcal{S}_{t,r}[m]^n$, where $0 \leqslant r \leqslant t$, and $r = 0$ if $m > t+1$.*

The proof of Theorem 1.3 uses a local stability analysis of the compression operator of Ahlswede and Khachatrian [2], and also the corresponding stability result for t-intersecting families in the p-biased hypercube obtained by Ellis, Keller and Lifshitz [6].

Notation

Throughout the paper, we write $[m] = \{1, \dots, m\}$. For any $x, y \in [m]^n$, we write $\mathsf{agr}(x, y) = |\{i \in [n] : x_i = y_i\}|$. We often identify a code $\mathcal{F} \subset [m]^n$ with its characteristic function $[m]^n \mapsto \{0,1\}$.

Given $x \in [m]^n$ and $R \subset [n]$, we define $x_R \in [m]^R$ by $(x_R)_i = x_i$. Given disjoint $R, R' \subset [n]$ and $a \in [m]^R$, $a' \in [m]^{R'}$, we sometimes denote their concatenation in $[m]^{R \cup R'}$ by $(x_R = a, x_{R'} = a')$.

Given $\alpha \in [m]^R$ for some $R \subset [n]$, we write $\mathcal{F}[\alpha] = \{x \in \mathcal{F} : x_R = \alpha\}$ and $\mathcal{F}(\alpha) = \{x \in [m]^{[n]\setminus R} : (x, \alpha) \in \mathcal{F}\}$. We also often denote $\mathcal{F}(\alpha)$ by $\mathcal{F}_{R\to\alpha}$.

For a coordinate $i \in [n]$ and symbol $a \in [m]$, we write $D_{i\to a}$ for the subcube having all $x \in [m]^n$ for which $x_i = a$; we will also refer to this as a 'dictator'. More generally, for $R \subset [n]$ and $a \in [m]^R$, we write $D_{R\to a} = \cap_{i\in R} D_{i\to a_i} = \{x \in [m]^n : x_R = a\}$.

Given $\mathcal{F} \subset [m]^n$ and $J \subset [n]$, we say that $\mathcal{F}$ is a J-junta if there is $\mathcal{A} \subset [m]^J$ such that $\mathcal{F} = \{x \in [m]^n : x_J \in \mathcal{A}\}$. When we do not wish to emphasise the set J itself, we instead refer to such families as $|J|$-juntas.

We will deal with various product domains $\Omega = \Omega_1 \times \dots \times \Omega_n$, mostly (but not only) with $\Omega = [m]^n$; we reserve μ to denote the uniform distribution over the domain under discussion (which will be clear from context). For any probability measure ν on Ω and $\mathcal{F} \subset \Omega$, we write $\nu(\mathcal{F}) = \sum_{x\in\mathcal{F}} \nu(x)$; similarly for $f : \Omega \to \mathbb{R}$, we write $\nu(f) = \mathbb{E}_{x\sim\nu} f(x) = \sum_{x\in\mathcal{F}} \nu(x) f(x)$.

We write $a \ll b$ to mean that there is some $a_0(b) > 0$ such that the following statement holds for $0 < a < a_0(b)$.

Part I: Small alphabets

Here we will prove our main result Theorem 1.1 when the alphabet size m is small, that is, t and m are fixed and $n > n_0(t, m)$ is large. This part of the paper will consist of three sections. In the next section, we prove Theorem 1.1 for fixed m, assuming three key steps of the proof (those described in the introduction). These steps are then proven as separate theorems in Sections 3 and 4.

We start with the junta approximation, for which the two key ingredients are (i) a regularity lemma, which approximately decomposes any code into pieces which are pseudorandom (in a sense to be made precise below), and (ii) a theorem of Mossel [24] on Markov chains hitting pseudorandom sets which implies that we can find a pair of vectors with any fixed agreement between any two pseudorandom families (of non-negligible measure).

In proving the stability version of the Ahlswede–Khachatrian anticode theorem, the first key observation is that for codes that are compressed (in a sense to be defined below), there is a natural transformation of the problem to the p-biased hypercube, where the stability theorem has already been proved by Ellis, Keller and Lifshitz [6]. This may at first not seem helpful for a general stability result, as compression destroys structure, but in fact we can make a local stability argument, that keeps control of the structure under gradual decompression, and thus deduce the general stability result.

For the bootstrapping step, the main ingredient is a 'cross disagreement' theorem, where given two families $\mathcal{F}$ and $\mathcal{G}$, we need to find $x \in \mathcal{F}$ and $y \in \mathcal{G}$ with $\mathrm{agr}(x, y) = 0$. We need this result in the unbalanced setting with $\mu(\mathcal{F}) = 1 - \alpha$ and $\mu(\mathcal{G}) = \beta$, where α and β are small, but α is large compared with β. The idea for overcoming this obstacle is to transform the problem via compressions to the setting of cross-intersecting families $\mathcal{F}'$ and $\mathcal{G}'$ in the p-biased hypercube, where $p = 1/m \leqslant 1/3$. We then move to the uniform (1/2-biased) measure, where by an isoperimetric lemma of Ellis, Keller and Lifshitz [7], the measure of the family corresponding to $\mathcal{G}$ becomes much larger, so that a trivial bound implies that $\mathcal{F}'$ and $\mathcal{G}'$ cannot be cross-intersecting.

2 | PROOF SUMMARY

In this section, we prove Theorem 1.1 for fixed m assuming the three theorems (junta approximation, anticode stability, bootstrapping) mentioned in the overview above, which we now state formally. The first theorem (junta approximation) proves Theorem 1.2 when J and n_0 can depend on m and replaces the conclusion $|\mathcal{F} \setminus \mathcal{J}| \leqslant \eta|\mathcal{J}|$ by $\mu(\mathcal{F} \setminus \mathcal{J}) \leqslant \eta$, which is an equivalent form when m is fixed; it will then remain to prove Theorem 1.2 for $m > m_0(t, \eta)$ sufficiently large (which we will do in Part II).

Theorem 2.1. *For every $\eta > 0$ and $t, m \in \mathbb{N}$ with $m \geqslant 3$, there are J and n_0 in $\mathbb{N}$ such that if $\mathcal{F} \subset [m]^n$ is a $(t-1)$-avoiding code with $n \geqslant n_0$, then there is a t-intersecting J-junta $\mathcal{J} \subset [m]^n$ such that $\mu(\mathcal{F} \setminus \mathcal{J}) \leqslant \eta$.*

The second theorem (anticode stability) is equivalent to Theorem 1.3 for fixed m and t, as we can bound $\mu(\mathcal{S}_{n,m,t})$ below by a constant.

Theorem 2.2. *For every $t \in \mathbb{N}$, $m \geqslant 3$ and $\varepsilon > 0$, there is $\delta > 0$ such that if $\mathcal{F} \subset [m]^n$ is t-intersecting with $\mu(\mathcal{F}) \geqslant \mu(\mathcal{S}_{n,m,t}) - \delta$, then $\mu(\mathcal{F} \setminus \mathcal{S}) \leqslant \varepsilon$ for some $\mathcal{S}$ which is isomorphic to some $\mathcal{S}_{n,m,t} = \mathcal{S}_{t,r}[m]^n$, where $0 \leqslant r \leqslant t$, and $r = 0$ if $m > t + 1$.*

The third theorem (bootstrapping) is an unbalanced cross disagreement theorem: it considers codes $\mathcal{G}, \mathcal{H} \subset [m]^n$ where $\mathcal{H}$ is small and $\mathcal{G}$ is almost complete, and finds $x \in \mathcal{F}$ and $y \in \mathcal{G}$ with $\mathrm{agr}(x, y) = 0$. We state it in a form that will also be useful later in the case that m is moderately large.

Theorem 2.3. *For every $t \in \mathbb{N}$ and $C > 0$, there is $\varepsilon_0 > 0$ such that if $0 < \varepsilon < \varepsilon_0$ and $\mathcal{G}, \mathcal{H} \subset [m]^n$ with $\mu(\mathcal{H}) = m^{-t}\varepsilon$ and $\mu(\mathcal{G}) > 1 - C\varepsilon$, then $\mathsf{agr}(x, y) = 0$ for some $x \in \mathcal{G}$ and $y \in \mathcal{H}$.*

Assuming these theorems, we now prove our main theorem for fixed m: the following is obtained from Theorem 1.1 by allowing n_0 to depend on m.

Theorem 2.4. *For all $t \in \mathbb{N}$ and $m \geqslant 3$, there is $n_0 \in \mathbb{N}$ such that if $\mathcal{F} \subset [m]^n$ is a $(t-1)$-avoiding code with $n \geqslant n_0$, then $|\mathcal{F}| \leqslant |\mathcal{S}_{n,m,t}|$, with equality only when $\mathcal{F}$ is isomorphic to a ball.*

Proof. Let $0 < n_0^{-1}, J^{-1} \ll \delta \ll \varepsilon \ll t^{-1}, m^{-1}$. Suppose $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding with $|\mathcal{F}| \geqslant |\mathcal{S}_{n,m,t}|$. By Theorem 2.1, there is a t-intersecting J-junta $\mathcal{J} \subset [m]^n$ such that $\mu(\mathcal{F} \setminus \mathcal{J}) \leqslant \delta$, and so $\mu(\mathcal{J} \cap \mathcal{F}) \geqslant \mu(\mathcal{F}) - \delta$. We have $\mu(\mathcal{J}) \geqslant \mu(\mathcal{F}) - \mu(\mathcal{F} \setminus \mathcal{J}) \geqslant \mu(\mathcal{S}_{n,m,t}) - \delta$. By Theorem 2.2 applied to $\mathcal{J}$, there is a copy $\mathcal{S}$ of $\mu(\mathcal{S}_{n,m,t})$ with $\mu(\mathcal{J} \setminus \mathcal{S}) \leqslant \varepsilon$. As $|\mathcal{F}| \geqslant |\mathcal{S}|$, we also have

$$\xi := \mu(\mathcal{S} \setminus \mathcal{F}) \leqslant \mu(\mathcal{F} \setminus \mathcal{S}).$$

Combining, we get

$$0 \leqslant \xi \leqslant \mu(\mathcal{F} \setminus \mathcal{S}) \leqslant \mu(\mathcal{F} \setminus \mathcal{J}) + \mu(\mathcal{J} \setminus \mathcal{S}) \leqslant \delta + \varepsilon.$$

Suppose for contradiction that $\xi > 0$. Without loss of generality, for some $r \leqslant t$, we can write

$$\mathcal{S} = \{x \in [m]^n : |\{i \in [t+2r] : x_i = 1\}| \geqslant t + r\}.$$

By averaging, there is $\alpha \in [m]^{[t+2r]}$ with $|\{i : \alpha_i = 1\}| < t + r$ such that $\mathcal{H} := \mathcal{F}_{[t+2r] \to \alpha}$ has $\mu(\mathcal{H}) \geqslant \mu(\mathcal{S} \setminus \mathcal{F}) = \xi$. We can fix $\beta \in [m]^{t+2r}$ with $|\{i : \beta_i = 1\}| \geqslant t + r$ such that $\mathsf{agr}(\alpha, \beta) = t - 1$. We have $\mu((\mathcal{S} \setminus \mathcal{F})_{[t+2r] \to \beta}) \leqslant m^{t+2r}\mu(\mathcal{S} \setminus \mathcal{F}) \leqslant m^{3t}\xi$, so $\mathcal{G} := \mathcal{F}_{[t+2r] \to \beta}$ has $\mu(\mathcal{G}) \geqslant 1 - m^{3t}\xi$.

By Theorem 2.3, with $C = m^{2t}$ and $m^t\xi$ in place of ε, we find $x \in \mathcal{G}$ and $y \in \mathcal{H}$ with $\mathsf{agr}(x, y) = 0$. However, this gives (α, y) and (β, x) in $\mathcal{F}$ with $\mathsf{agr}((\alpha, y), (\beta, x)) = t - 1$, which is a contradiction. □

3 | JUNTA APPROXIMATION

In this section, we prove the junta approximation theorem for fixed m, that is, Theorem 2.1. Our first ingredient is a regularity lemma, showing that any code can be approximately decomposed into a constant number of pieces, each of which is pseudorandom, in the sense that restrictions of constant size do not significantly affect the measure. This regularity lemma is similar in spirit to that in [6, Theorem 1.7]; we refer the reader to Section 1.2 of their paper for discussion how such results are related to the large literature on regularity lemmas in Combinatorics.

The second ingredient is a result of Mossel [24] on Markov chains hitting pseudorandom sets, which implies that any two pseudorandom codes $\mathcal{F}, \mathcal{G} \subset [m]^n$ of non-negligible measure cannot be cross intersecting, that is, we can find a 'disagreement' $(x, y) \in \mathcal{F} \times \mathcal{G}$ with $\mathsf{agr}(x, y) = 0$. If $\mathcal{F}$ is $(t-1)$-avoiding, this will imply $\mathsf{agr}(\alpha, \beta) \geqslant t$ for any pieces $\mathcal{F}_{T \to \alpha}$, $\mathcal{F}_{T \to \beta}$ of the regularity decomposition of $\mathcal{F}$ that are pseudorandom and of non-negligible measure. Indeed, if we had $\mathsf{agr}(\alpha, \beta) = t - 1 - s$ with $s \geqslant 0$, then we could arbitrarily fix a further restriction $S \to \gamma$ with

$|S| = s$ to obtain pseudorandom families $\mathcal{F}_{(T,S)\to(\alpha,\gamma)}$, $\mathcal{F}_{(T,S)\to(\beta,\gamma)}$ that are cross intersecting, which is impossible. Here we are implicitly using the (important) fact that pseudorandomness is preserved by constant size restrictions.

3.1 | The pseudorandom code regularity lemma

In this subsection, we prove a regularity lemma which approximately decomposes any code into pieces that are pseudorandom in the sense of the following definition.

Definition 3.1. We say $\mathcal{F} \subset [m]^n$ is (r, ε)-pseudorandom if for any $R \subset [n]$ with $|R| \leqslant r$ and $a \in [m]^R$, we have $|\mu(\mathcal{F}_{R\to a}) - \mu(\mathcal{F})| \leqslant \varepsilon$.

Lemma 3.2. *For any $r, m \in \mathbb{N}$ and $\varepsilon, \delta > 0$, there is $D \in \mathbb{N}$ such that for any $\mathcal{F} \subset [m]^n$ with $n \geqslant D$, there is $T \subset [n]$ with $|T| \leqslant D$ such that $\Pr_{\mathbf{a}\in[m]^T}[\mathcal{F}_{T\to\mathbf{a}}$ is not (r, ε)-pseudorandom$] \leqslant \delta$.*

Proof. We construct T iteratively. Starting with $T = \emptyset$, we consider at each step the set A of $a \in [m]^T$ for which $\mathcal{F}_{T\to a}$ is not (r, ε)-pseudorandom. For any $a \in A$, we fix $b(a) \in [m]^{R_a}$ for some $R_a \subset [n]$ with $|R_a| \leqslant r$ such that $|\mu(\mathcal{F}_{(T,R_a)\to(a,b(a))}) - \mu(\mathcal{F}_{T\to\alpha})| > \varepsilon$. If $\mu(A) \leqslant \delta$ we are done; otherwise, we replace T by $T_{\mathsf{new}} = T \cup R$ where $R = \bigcup_{a\in A} R_a$ and iterate.

We will argue that this process stops with $|T|$ bounded by some function depending on m, r, δ and ε, but not on n. To do so, we apply a standard 'energy increment' argument to the mean-square density

$$E(T) = \mathop{\mathbb{E}}_{\mathbf{a}\in[m]^T} \left[\mu(\mathcal{F}_{T\to\mathbf{a}})^2\right].$$

Clearly, $E(T) \leqslant 1$ for any $T \subset [n]$, and $E(T_1) \leqslant E(T_2)$ whenever $T_1 \subset T_2$ by Cauchy–Schwarz.

We will show that $E(T)$ increases significantly at each step of the process. Indeed, comparing $E(T_{\mathsf{new}})$ and $E(T)$ term by term, we have

$$E(T_{\mathsf{new}}) - E(T) = \mathop{\mathbb{E}}_{\mathbf{a}\in[m]^T}\left[\mathop{\mathbb{E}}_{\mathbf{b}\in[m]^R}\left[\mu(\mathcal{F}_{(T,R)\to(\mathbf{a},\mathbf{b})})\right]^2 - \mu(\mathcal{F}_{T\to\mathbf{a}})^2\right] = \mathop{\mathbb{E}}_{\mathbf{a}\in[m]^T}[\operatorname{Var} Z_{\mathbf{a}}],$$

where we consider $Z_a(\mathbf{b}) = \mu(\mathcal{F}_{(T,R)\to(a,\mathbf{b})})$ as a random variable determined by the random choice of $\mathbf{b} \in [m]^R$. We have $\operatorname{Var} Z_a \geqslant 0$ for all a, and for any $a \in A$, we have $\operatorname{Var} Z_a \geqslant m^{-|R_a|}\varepsilon^2 \geqslant m^{-r}\varepsilon^2$ in light of the restriction $R_a \to b(a)$. Therefore, $E(T_{\mathsf{new}}) \geqslant E(T) + \mu(A)m^{-r}\varepsilon^2 \geqslant E(T) + \delta m^{-r}\varepsilon^2$.

In other words, as long as the process does not terminate, the energy function increases by at least $\delta m^{-r}\varepsilon^2$. As the energy is always at most 1, the process terminates after at most $m^r/\delta\varepsilon^2$ steps. Each restriction adds at most r new variables to T, so in each step, $|T_{\mathsf{new}}| \leqslant 2^{|T|} \cdot r$, and so, the final size of T is bounded by some function of m, r, δ and ε. $\square$

3.2 | Markov chains hitting pseudorandom sets

In this subsection, we discuss a special case of a result of Mossel [24] needed for the proof of our junta approximation theorem for small alphabets, which can be formulated in terms of Markov chains hitting pseudorandom sets. We start by summarising some properties of Markov chains (see [21] for an introduction). We will consider finite Markov chains, that is, a sequence of random variables $(X_i)_{i\geqslant 0}$ taking values in a state space S (some finite set) described by a transition matrix T with rows and columns indexed by S, where for any event E determined by $(X_0, \dots, X_i)$ with $X_i = x$, we have $\mathbb{P}(X_{i+1} = y \mid X_i = x) = T_{xy}$. We also view T as an averaging operator on functions $f : S \to \mathbb{R}$, corresponding to matrix multiplication when we view f as a vector in $\mathbb{R}^S$: we have $(Tf)(x) = \mathbb{E}[f(X_1) \mid X_0 = x] = \sum_y T_{xy} f(y) = (Tf)_x$.

We will suppose that T is irreducible (for any $x, y \in S$, there is some $k \in \mathbb{N}$ with $T^k_{xy} > 0$), so there is a unique stationary distribution (a probability distribution ν on S such that $\nu T = \nu$). The stationary chain is obtained by letting X_0 have distibution ν, and then each X_i has distribution ν. In the stationary chain, we have $\mathbb{P}(X_0 = a, X_1 = b) = P_{ab} := \nu_a T_{ab}$. We say that T is reversible if P is symmetric, that is, $P_{ab} = P_{ba}$ for all $a, b \in S$ (the name corresponds to the observation that the distribution of the stationary chain is invariant under time reversal).

When T is reversible, it defines a self-adjoint operator on $L^2(S, \nu)$, that is, functions $f : S \to \mathbb{R}$ with the inner product $\langle f, g\rangle = \sum_x \nu_x f(x) g(x)$, so $L^2(S, \nu)$ has an orthonormal basis B of eigenfunctions of T. We can write any $f \in L^2(S, \nu)$ in the form $f = \sum_{b\in B} c_b b$, and then $\mathbb{E} f^2 = \langle f, f\rangle = \sum_{b\in B} c_b^2$. The largest eigenvalue is 1, and the corresponding eigenspace consists of constant functions on S. If $Tf = \lambda f$ with $\lambda \neq 1$, then $\mathbb{E} f := \mathbb{E}_{x\sim\nu} f(x) = \sum_x \nu_x f(x) = \langle f, 1\rangle = 0$. The absolute spectral gap λ_* is the minimum value of $1 - |\lambda|$ over all eigenvalues $\lambda \neq 1$; equivalently,

$$(1 - \lambda_*)^2 = \sup\{\mathbb{E}(Tf)^2 : \mathbb{E} f = 0, \mathbb{E} f^2 = 1\}.$$

Now we describe a special case of [24, Theorem 4.4], and for that, we require a basic setup. Let T be a reversible, irreducible Markov chain acting on $S = [m]$, and consider its tensor power $T^{\otimes n}$ acting on $\Omega = [m]^n$ independently in each coordinate, that is, with transition matrix $T^{\otimes n}_{xy} = \prod_{i=1}^n T_{x_i y_i}$. In essence, [24, Theorem 4.4] asserts that if T has a constant spectral gap, and we have pseudorandom codes $\mathcal{F}, \mathcal{G} \subset [m]^n$ of noticeable measure, then sampling consecutive random states x, y of the stationary chain for $T^{\otimes n}$, we have that $x \in \mathcal{F}$, $y \in \mathcal{G}$ with significant probability.

Theorem 3.3. *Let T be a reversible Markov chain on $[m]$ with absolute spectral gap $\lambda_* > 0$. Let ν denote the stationary measure of $T^{\otimes n}$ and x and y be consecutive random states of the stationary chain. Then for any $\mu > 0$, there are $\varepsilon, c > 0$ and $r \in \mathbb{N}$ such that if $\mathcal{F}, \mathcal{G} \subset [m]^n$ are (r, ε)-pseudorandom with $\nu(\mathcal{F}), \nu(\mathcal{G}) > \mu$, then $\mathbb{P}(x \in \mathcal{F}, y \in \mathcal{G}) > c$.*

For convenience of the reader, we outline below the (standard) derivation of Theorem 3.3 from existing results in the literature.

Deriving Theorem 3.3 from [24, Theorem 4.4]

Let $B = \{b_1, \dots, b_m\}$ be an orthonormal basis for $L^2(S, \nu)$ consisting of eigenvectors of T. We take b_1 to be the trivial eigenvector, that is, $b_1(s) = 1$ for all $s \in S$, which has eigenvalue 1. We remark that by the spectral gap of T, it follows that the eigenvalue of each b_j for $j \neq 1$ is at most $1 - \lambda_*$. We will view each b_i as a random variable on (S, ν), and in this language, we have that $\mathbb{E} b_i b_j = 1_{i=j}$. Using the basis B, we may find a basis for $L^2(S^n, \nu^{\otimes n})$ by tensorizing. Namely, for each $i \in [n]$, we take an independent copy of B, say $b^i = (b^i_j : j \in [m])$, and then our basis is $\boldsymbol{b} = (\boldsymbol{b}_{j_1,\dots,j_n})_{j_1,\dots,j_n\in[m]}$, where $\boldsymbol{b}_{j_1,\dots,j_n} = \prod_{i=1}^n b^i_{j_i}$. We can thus represent any function on Ω as a multi-linear polynomial $P(\boldsymbol{b}) = \sum_\alpha c_\alpha b^\alpha$ where α ranges over $[m]^n$ and $b^\alpha := \prod_i b^i_{\alpha_i}$.

This above view allows us to extend the definition of P to $\mathbb{R}^{mn}$. A technical point to note, however, is that even if our original function P was bounded on $[m]^n$ (in our case, it is even Boolean valued), the extension to $\mathbb{R}^{mn}$ may not be bounded. For this reason, one first applies a small noise on the function P, that is, considers $Q(x) = T_{1-\eta}P(x) = \mathbb{E}_{x'\sim_{1-\eta}x}[P(x')]$ where for each $i \in [n]$ independently, $x'_i = x_i$ with probability $1-\eta$ and otherwise x'_i is resampled according to ν ($\eta > 0$ is to be thought of as a small constant, much smaller than the spectral gap λ_* of T), and then truncates it. Namely, consider the multi-linear extension of Q, $Q(b)$ as defined above, and let $\tilde{P}(b) = Q(b)$ if $0 \leqslant Q(b) \leqslant 1$, $\tilde{P}(b) = 1$ if $Q(b) > 1$, and otherwise $\tilde{P}(b) = 0$.

Let x and y be sampled as consecutive random states of the stationary chain for $T^{\otimes n}$, and let $f(\boldsymbol{b}(x)) = 1_{x\in\mathcal{F}}$, $g(\boldsymbol{b}(y)) = 1_{y\in\mathcal{G}}$. Our goal is thus to prove a lower bound on $\mathbb{E}_{x,y}[f(\boldsymbol{b}(x))g(\boldsymbol{b}(y))]$. The invariance principles of [23–25] allow one to establish non-trivial lower bounds on this quantity by considering its 'analogue in Gaussian space', provided that f, g are sufficiently random-like.

To be more precise, let us first consider $\boldsymbol{b}_{j_1,\dots,j_n}(x)$ and $\boldsymbol{b}_{j'_1,\dots,j'_n}(y)$ where x and y are sampled as consecutive random states of the stationary chain for $T^{\otimes n}$. Thus, $\mathbb{E}b^i_{j_i}(x)b^{i'}_{j'_{i'}}(y)$ is zero unless $i = i'$ and $j_i = j'_{i'}$, and then, it is equal to the eigenvalue λ_{j_i} such that $Tb_{j_i} = \lambda_{j_i}b_{j_i}$. We now wish to define the Gaussian analog of $\boldsymbol{b}_{j_1,\dots,j_n}(x)$ and $\boldsymbol{b}_{j_1,\dots,j_n}(y)$. Let $Z = \{z_1, \dots, z_m, z'_1, \dots, z'_m\}$ be Gaussian variables with the same covariance matrix. Namely, we take $z_1 = z'_1 = 1$, and $z_2, \dots, z_m$ and $z'_2, \dots, z'_m$ are jointedly distributed standard Gaussian random variables such that $z_2, \dots, z_m$ are independent, $z'_2, \dots, z'_m$ are independent and $\mathbb{E}[z_j z'_{j'}] = \mathbb{E}[b_j(x)b_{j'}(y)] = \lambda_j 1_{j=j'}$. We take n independent copies of Z, $Z^i = \{{z^i}_1, \dots, {z^i}_m, {z^{i}}'_1, \dots, {z^{i}}'_m\}$, and then define $\boldsymbol{z}_{j_1,\dots,j_n} = \prod_{i=1}^n {\boldsymbol{z}^i}_{j_i}$ and $\boldsymbol{z}'_{j_1,\dots,j_n} = \prod_{i=1}^n {\boldsymbol{z}^{i}}'_{j_i}$. The random variables $\boldsymbol{z}_{j_1,\dots,j_n}$, $\boldsymbol{z}'_{j_1,\dots,j_n}$ are to be thought of as the Gaussian analogs of $\boldsymbol{b}_{j_1,\dots,j_n}(x)$ and $\boldsymbol{b}_{j_1,\dots,j_n}(y)$.

Building on [25], Mossel [23] showed that for $f, g : [m]^n \to [0,1]$ with 'small enough influences',† one has $\mathbb{E}[f(\boldsymbol{b}(x)) \cdot g(\boldsymbol{b}(y))]$ is very close $\mathbb{E}[\tilde{f}(\boldsymbol{z})\tilde{g}(\boldsymbol{z}')]$. The arguments in [24] establish the same statement with the more relaxed condition that f and g are (r,ε)-pseudorandom (the term 'resilient' is used therein). More precisely, Mossel showed that for all $\delta > 0$, there are $r \in \mathbb{N}$ and $\varepsilon > 0$ (also depending on m and the spectral gap λ_*, which are thought of as constants), such that $|\mathbb{E}[f(\boldsymbol{b}(x)) \cdot g(\boldsymbol{b}(y))] - \mathbb{E}[\tilde{f}(\boldsymbol{z})\tilde{g}(\boldsymbol{z}')]| \leqslant \delta$.

For $\tilde{f}$, the fact that f has averages at least μ implies, by the invariance principle (i.e. the above with $g = 1$), that $\tilde{f}$ has average at least $\mu/2$; similarly, the average of $\tilde{g}$ is at least $\mu/2$. Thus, $\mathbb{E}[\tilde{f}(\boldsymbol{z})\tilde{g}(\boldsymbol{z}')] > c(\lambda_*, \mu) > 0$ by reverse hypercontractivity (see [15, Theorem A.78], e.g.), and as this is close to $\mathbb{E}[f(\boldsymbol{b}(x)) \cdot g(\boldsymbol{b}(y))]$, we get that $\mathbb{E}[f(\boldsymbol{b}(x)) \cdot g(\boldsymbol{b}(y))] \geqslant c/2$, establishing Theorem 3.3.

The following result is an immediate consequence of Theorem 3.3, applied with the Markov chain T on $[m]$ which at each step moves to a uniformly random state different from the current state (note that $\lambda_* > 0$ when $m \geqslant 3$, but this fails for $m = 2$).

Theorem 3.4. *For every $m \geqslant 3$ and $\mu > 0$, there are $\varepsilon, c > 0$ and $r \in \mathbb{N}$ such that if $\mathcal{F}, \mathcal{G} \subset [m]^n$ are (r,ε)-pseudorandom with $\mu(\mathcal{F}), \mu(\mathcal{G}) > \mu$ and (x,y) is a uniformly random pair in $[m]^n \times [m]^n$ with $\mathsf{agr}(x,y) = 0$, then $\mathbb{P}(x \in \mathcal{F}, y \in \mathcal{G}) > c$; in particular, $\mathsf{agr}(x,y) = 0$ for some $(x,y) \in \mathcal{F} \times \mathcal{G}$.*

† We omit the definition of 'influences' for now, as we do not need it here, but it will reappear later in a more general context when we discuss our theory of global hypercontractivity.

3.3 | Approximation by junta

We conclude this section by proving Theorem 2.1.

Proof of Theorem 2.1. Let $t, m \in \mathbb{N}$ with $m \geqslant 3$ and $\eta > 0$, fix $0 \ll n_0^{-1} \ll D^{-1} \ll r^{-1}, \varepsilon \ll \eta, t^{-1}, m^{-1}$ and suppose $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding. By Lemma 3.2, we find $T \subset [n]$ with $|T| \leqslant D$ such that

$$\Pr_{\mathbf{a} \in [m]^T} [\mathcal{F}_{T \to \mathbf{a}} \text{ is not } (r, \varepsilon)\text{-pseudorandom}] \leqslant \eta/2.$$

We will show that the required conclusions of the theorem hold for the junta $\mathcal{J} = \{x \in [m]^n \mid x_T \in J\}$, where

$$J = \left\{ \alpha \in [m]^T \,\middle|\, \mathcal{F}_{T \to \alpha} \text{ is } (r, \varepsilon/2)\text{-pseudorandom and } \mu(\mathcal{F}_{T \to \alpha}) \geqslant \eta/2 \right\},$$

that is, that $\mathcal{J}$ is t-intersecting (equivalently, J is t-intersecting) and $\mathcal{F}$ is approximately contained in $\mathcal{J}$.

To see that J is t-intersecting, suppose for contradiction, we have $\alpha_1, \alpha_2 \in J$ with $\mathsf{agr}(\alpha_1, \alpha_2) = t - 1 - s$ with $s \geqslant 0$. Fix $S \subset [n] \setminus T$ of size s and $x \in [m]^S$ arbitrarily, and consider the families

$$\mathcal{G}_i = \left\{ w \in [m]^{[n] \setminus (T \cup S)} \,\middle|\, (\alpha_i, x, w) \in \mathcal{F} \right\}$$

for $i = 1, 2$. By definition of J, both $\mu(\mathcal{G}_i) \geqslant \mu(\mathcal{F}_{\alpha_i}) - \varepsilon/2 \geqslant \eta/3$ and $\mathcal{G}_i$ is $(r - t, \varepsilon)$-pseudorandom. By Theorem 3.4, we find $(w_1, w_2) \in \mathcal{G}_1 \times \mathcal{G}_2$ with $\mathsf{agr}(w_1, w_2) = 0$. However, this gives (α_i, x, w_i) for $i = 1, 2$ in $\mathcal{F}$ with agreement $t - 1$, which is a contradiction.

It remains to bound $\mu(\mathcal{F} \setminus \mathcal{J}) = \sum_{\alpha \notin J} m^{-|T|} \mu(\mathcal{F}_{T \to \alpha})$. We partition $[m]^T \setminus J$ into (B_1, B_2) where B_1 contains those $\alpha \in [m]^T \setminus J$ with $\mu(\mathcal{F}_{T \to \alpha}) < \eta/2$, and $B_2 = [m]^T \setminus (B_1 \cup J)$. Clearly, the contribution to the sum from $\alpha \in B_1$ is at most $\eta/2$. For $\alpha \in B_2$, we note that $\mathcal{F}_{T \to \alpha}$ is not $(r, \varepsilon/2)$-pseudorandom by definition of J, so $\sum_{\alpha \in B_2} m^{-|T|} \mu(\mathcal{F}_{T \to \alpha}) \leqslant \mu(B_2) < \eta/2$ by choice of T. Thus, $\mu(\mathcal{F} \setminus \mathcal{J}) < \eta$. □

4 | COMPRESSION, STABILITY AND BOOTSTRAPPING

In this section, we prove the anticode stability theorem for fixed m, that is, Theorem 2.2, and the bootstrapping result Theorem 2.3. Both rely on a compression procedure, introduced by Ahlswede and Khachatrian [2], which modifies any code in such a way to use some symbol (say 1) 'as much as possible', while maintaining its size and not reducing its minimum intersection size.

In the first subsection, we will formally define compression and prove some of its well-known properties. In the second subsection, we prove the stability result for compressed codes, by reducing it to the corresponding stability result for t-intersecting families in the biased hypercube obtained by Ellis, Keller and Lifshitz [6]. We deduce Theorem 2.2 in the third subsection, via a decompression argument, in which we reverse the compressions while keeping control of structure via a local stability argument. In the final subsection, we prove Theorem 2.3, by using compressions to reformulate the problem in terms of cross-intersecting families in the biased hypercube.

4.1 Compression

For any $i \in [n]$ and $j \in [m]$, we define the compression operator $T_{i,j} : [m]^n \to [m]^n$ that replaces j by 1 in coordinate i if possible, that is, for $x \in [m]^n$, we let $T_{i,j}(x) = y \in [m]^n$ where $y_r = x_r$ for all $r \neq i$, and $y_i = x_i$ if $x_i \neq j$ or $y_i = 1$ if $x_i = j$. We also define a compression operator, also denoted as $T_{i,j}$, on codes, that replaces any vector x by $T_{i,j}(x)$ unless the latter is already present, that is,

$$T_{i,j}(\mathcal{F}) = \left\{ x \mid T_{i,j}(x) \in \mathcal{F} \right\} \cup \left\{ T_{i,j}(x) \,\middle|\, x \in \mathcal{F} \right\}.$$

We also define $T_i = T_{i,2} \circ T_{i,3} \circ \dots \circ T_{i,m}$ for any $i \in n$, and $T = T_1 \circ T_2 \circ \dots \circ T_n$. One can think of T as trying to set as many coordinates as possible equal to 1. We call $\mathcal{F} \subset [m]^n$ compressed if $T(\mathcal{F}) = \mathcal{F}$.

We need the following well-known facts about these compression operators.

Fact 4.1. *Let $\mathcal{F}, \mathcal{G} \subset [m]^n$, $i \in [n]$ and $j \in [m]$.*

(1) *We have $\mu(T(\mathcal{F})) = \mu(T_{i,j}(\mathcal{F})) = \mu(\mathcal{F})$.*
(2) *The family $T(\mathcal{F})$ is compressed.*
(3) *If $\mathcal{F}, \mathcal{G}$ are cross t-intersecting, then so are $T_{i,j}(\mathcal{F})$ and $T_{i,j}(\mathcal{G})$, and so are $T(\mathcal{F})$ and $T(\mathcal{G})$. Furthermore, any $x \in T(\mathcal{F}), y \in T(\mathcal{G})$ have at least t common coordinates equal to 1.*

Proof. We begin with the first item. Assume without loss of generality that $i = 1$. To see that $\mu(T_{i,j}(\mathcal{F})) = \mu(\mathcal{F})$, we consider any $x \in [m]^{n-1}$, note that vectors (a, x) with $a \in [m] \setminus \{1, j\}$ are unaffected by $T_{i,j}$, and that $T_{i,j}(\mathcal{F})$ and $\mathcal{F}$ contain the same number of elements of $\{(1, x), (j, x)\}$. By iterating, we deduce $\mu(T(\mathcal{F})) = \mu(\mathcal{F})$.

Now we address the second item. Say that $\mathcal{F}$ is i-compressed for some i if whenever $x \in T_i\mathcal{F}$ has $x_i = j \neq 1$, we have $T_{i,j}x \in \mathcal{F}$. We first claim that $T_i\mathcal{F}$ is i-compressed for each i. Indeed, as $T_i\mathcal{F} = T_{i,2} \circ \dots T_{i,j} \circ \dots \circ T_{i,m}\mathcal{F}$, it follows that either $T_{i,j}x \in \mathcal{F}$ and then we are done, or else there is $j' < j$ such that $\mathcal{F}$ contains the point y that only differs from x in coordinate i and has $y_i = j'$. In that case, $T_{i,j'}y = T_{i,j}x$ would be in $T_i\mathcal{F}$, so again the claim holds.

Next, we argue that if $\mathcal{F}$ is i-compressed, then $T_{i'}\mathcal{F}$ is still i-compressed. To see this, we consider any $x \in T_i \circ T_{i'}\mathcal{F}$ with $x_i = j \neq 1$ and show that $T_{i,j}x \in T_{i'}\mathcal{F}$. We note that $x \in T_{i'}\mathcal{F}$ and consider two cases.

(a) If $x_{i'} = j' \neq 1$, then x must be in $\mathcal{F}$ and as $\mathcal{F}$ is i-compressed, we get that $T_{i,j}x \in \mathcal{F}$; also, as $x \in T_{i'}\mathcal{F}$, we get that $T_{i',j'}x \in \mathcal{F}$, and as $\mathcal{F}$ is i-compressed, it follows that $T_{i,j} \circ T_{i',j'}x \in \mathcal{F}$. Letting $y = T_{i,j}x$, we see that y and $T_{i',j'}y = T_{i,j} \circ T_{i',j'}x$ are both in $\mathcal{F}$, and so $y \in T_{i'}\mathcal{F}$.
(b) If $x_{i'} = 1$, then there is $j' \geqslant 1$ such that for the point y that is the same as x on all coordinates except for $y_i = j'$, we have $y \in \mathcal{F}$. As $\mathcal{F}$ is i-compressed, we get that $T_{i,j}y \in \mathcal{F}$, and so $T_{i',j'} \circ T_{i,j}y \in T_{i'}\mathcal{F}$. Noting that $T_{i',j'} \circ T_{i,j}y = T_{i,j}x$, we conclude that $T_{i,j}x \in T_{i'}\mathcal{F}$.

It follows that the family $T\mathcal{F}$ is i-compressed for all $i \in [n]$. To conclude, we claim that $T_i \circ T(\mathcal{F}) = T(\mathcal{F})$, which readily implies the second item. Indeed, if $x \in T_i \circ T(\mathcal{F})$, then letting $j = x_i$, if $j \neq 1$, then $x \in T(\mathcal{F})$. If $j = 1$, then there is $j' \geqslant 1$ such that $T(\mathcal{F})$ contains the point y that is the same as x except for $y_i = j'$. As $T(\mathcal{F})$ is i-compressed, $x = T_{i,j'}y \in T(\mathcal{F})$. Overall, $T_i \circ T(\mathcal{F}) \subseteq T(\mathcal{F})$. Combining with the first item, it follows that the two families are equal.

We now move on to the third item, and suppose for contradiction that $\mathcal{F}, \mathcal{G}$ are cross t-intersecting but $T_{1,j}(\mathcal{F})$ and $T_{1,j}(\mathcal{G})$ are not. Then there are $(a, x) \in T_{i,j}(\mathcal{F})$ and $(b, y) \in T_{i,j}(\mathcal{G})$

with $\mathrm{agr}((a,x),(b,y)) < t$. As $\mathcal{F},\mathcal{G}$ are cross t-intersecting, we cannot have both $(a,x) \in \mathcal{F}$ and $(b,y) \in \mathcal{G}$, so without loss of generality, $(a,x) = (1,x)$ was obtained from $(j,x) \in \mathcal{F}$. We must have $(b,y) \in \mathcal{G}$, as otherwise $(b,y) = (1,y)$ was obtained from $(j,y) \in \mathcal{G}$, but then $(j,x) \in \mathcal{F}$ and $(j,y) \in \mathcal{G}$ with $\mathrm{agr}((j,x),(j,y)) = \mathrm{agr}((1,x),(1,y)) < t$, contradiction. As $(j,x) \in \mathcal{F}$ and $(b,y) \in \mathcal{G}$, we have $\mathrm{agr}((j,x),(b,y)) \geqslant t$, so $b = j$. As $(b,y) \in T_{i,j}(\mathcal{G})$, we must have $(1,y) \in \mathcal{G}$. But now $\mathrm{agr}((j,x),(1,y)) < t$ gives a contradiction. Thus, $T_{1,j}(\mathcal{F})$ and $T_{1,j}(\mathcal{G})$ are cross t-intersecting. By iterating, so are $T(\mathcal{F})$ and $T(\mathcal{G})$.

Finally, suppose for contradiction that $x \in T(\mathcal{F}), y \in T(\mathcal{G})$ have fewer than t common coordinates equal to 1. Let x' be obtained from x by setting $x'_i = 1$ if $x_i = y_i \neq 1$ or $x'_i = x_i$ otherwise. Then $x' \in T(\mathcal{F})$ but x' and y only agree on coordinates i with $x_i = y_i = 1$, which contradicts $T(\mathcal{F})$ and $T(\mathcal{G})$ being cross t-intersecting. □

Next, we will define a transformation from compressed codes in $[m]^n$ to monotone[†] families in the cube $\{0,1\}^n$ that preserves minimum (cross) intersection size, and does not decrease the measure when we adopt the p-biased measure on the cube with $p = 1/m$ (as we will do throughout this section).

Definition 4.2. We define $h\colon [m] \to \{0,1\}$ by $h(1) = 1$ and $h(a) = 0$ for all $a \neq 1$, and $h^{\otimes n}\colon [m]^n \to \{0,1\}^n$ by $h^{\otimes n}(x) = (h(x_1),\dots,h(x_n))$. For any $\mathcal{F} \subset [m]^n$, we let $\widetilde{\mathcal{F}} = h^{\otimes n}(\mathcal{F}) \subset \{0,1\}^n$.

Fact 4.3. *Suppose $\mathcal{F},\mathcal{G} \subset [m]^n$ are compressed.*

(1) *The family $\widetilde{\mathcal{F}}$ is monotone and $\mu_p(\widetilde{\mathcal{F}}) \geqslant \mu(\mathcal{F})$.*
(2) *If $\mathcal{F}$ is t-intersecting, then so is $\widetilde{\mathcal{F}}$.*
(3) *If $\mathcal{F},\mathcal{G}$ are cross t-intersecting, then so are $\widetilde{\mathcal{F}},\widetilde{\mathcal{G}}$.*

Proof. The intersection statements are immediate from the final part of Fact 4.1. For monotonicity, consider any $\tilde{x} \in \widetilde{\mathcal{F}}$ and $\tilde{y} \geqslant \tilde{x}$. Fix $x \in \mathcal{F}$ with $h^{\otimes n}(x) = \tilde{x}$, that is, $x_i = 1$ if and only if $\tilde{x}_i = 1$. Define $y \in [m]^n$ by $y_i = 1$ if $\tilde{y}_i = 1 \neq \tilde{x}_i$ or $y_i = x_i$ otherwise. Then $y \in \mathcal{F}$, as $\mathcal{F}$ is compressed, and $h^{\otimes n}(y) = \tilde{y}$, so $\tilde{y} \in \widetilde{\mathcal{F}}$.

To show $\mu_p(\widetilde{\mathcal{F}}) \geqslant \mu(\mathcal{F})$, we consider intermediate product spaces $\{0,1\}^r \times [m]^{n-r}$ with the measure $\nu_r = \mu_p^r \times \mu$, and intermediate families $\mathcal{F}_r = (h^{\otimes r} \otimes I^{\otimes n-r})(T(\mathcal{F}))$ for any $r \geqslant 0$. It suffices to show $\nu_{r+1}(\mathcal{F}_{r+1}) \geqslant \nu_r(\mathcal{F}_r)$ for any $r \geqslant 0$. We can write

$$\nu_{r+1}(\mathcal{F}_{r+1}) - \nu_r(\mathcal{F}_r) = \sum_{x\in\{0,1\}^r, y\in[m]^{n-r-1}} \mu_p(x) m^{-(n-r-1)}(\mu_p(B_{x,y,r}) - |A_{x,y,r}|/m),$$

where for $0 \leqslant r \leqslant n$ and $x \in \{0,1\}^r$, $y \in [m]^{n-r-1}$, we define

$$A_{x,y,r} = \left\{ a \in [m] \,\middle|\, (x,a,y) \in \mathcal{F}_r \right\}, \qquad B_{x,y,r} = \left\{ a \in \{0,1\} \,\middle|\, (x,a,y) \in \mathcal{F}_{r+1} \right\}.$$

Thus, it suffices to show $\mu_p(B_{x,y,r}) \geqslant \frac{|A_{x,y,r}|}{m}$ for all x, y. To see this, suppose first that $|A_{x,y,r}| = 1$. As $\mathcal{F}$ is compressed, we have $A_{x,y,r} = \{1\}$, so $B_{x,y,r} = h(A_{x,y,r}) = \{1\}$ and $\mu_p(B_{x,y,r}) = p =$

[†] We call $\mathcal{A} \subset \{0,1\}^n$ monotone if $y \in \mathcal{G}$ whenever $x \in \mathcal{G}$ and $x \leqslant y$ coordinatewise.

$|A_{x,y,r}|/m$. Otherwise, if $|A_{x,y,r}| \geqslant 2$, we have $B_{x,y,r} = h(A_{x,y,r}) = \{0,1\}$, so $\mu_p(B_{x,y,r}) = 1 \geqslant |A_{x,y,r}|/m$. □

4.2 | Stability when compressed

In this subsection, we prove Theorem 2.2 for compressed families, using the corresponding stability result for t-intersecting families in the biased hypercube obtained by Ellis, Keller and Lifshitz [6], which we start by stating. Given n, p, t, let $\mathcal{S}_{n,p,t}$ denote a family $\mathcal{S}_{t,r}\{0,1\}^n \subset \{0,1\}^n$ with the largest p-biased measure, where $r = 0, 1, \dots, t-1$ and

$$\mathcal{S}_{t,r}\{0,1\}^n = \{x \in \{0,1\}^n : |\{i \in [t+2r] : x_i = 1\}| \geqslant t+r\}.$$

The following is implied by [6, Theorem 1.10]. †

Theorem 4.4. *For every $t \in \mathbb{N}$, $\zeta > 0$ and $\varepsilon > 0$, there is $\delta > 0$ such that if $\mathcal{F} \subset \{0,1\}^n$ is t-intersecting, $\zeta \leqslant p \leqslant \frac{1}{2} - \zeta$ and $\mu_p(\mathcal{F}) \geqslant (1-\delta)\mu(\mathcal{S}_{n,p,t})$, then $\mu_p(\mathcal{F} \setminus \mathcal{S}) \leqslant \varepsilon\mu(\mathcal{S})$ for some copy $\mathcal{S}$ of $\mathcal{S}_{n,p,t} = \mathcal{S}_{t,r}\{0,1\}^n$, where $0 \leqslant r \leqslant t$ if $p \leqslant 1/3$, and $r = 0$ if $p \leqslant \frac{1}{t+1} - \zeta$.*

Using Theorem 4.4, we can prove a weaker version of Theorem 2.2, with the additional assumption that $\mathcal{F}$ is compressed. This version will be used in the next subsection to prove Theorem 2.2 as stated.

Claim 4.5. For every $t \in \mathbb{N}$, $m \geqslant 3$ and $\varepsilon > 0$, there is $\delta > 0$ such that if $\mathcal{F} \subset [m]^n$ is compressed and t-intersecting with $\mu(\mathcal{F}) \geqslant (1-\delta)\mu(\mathcal{S}_{n,m,t})$, then $\mu(\mathcal{F} \setminus \mathcal{S}) \leqslant \varepsilon\mu(\mathcal{S})$ for some copy $\mathcal{S}$ of $\mathcal{S}_{n,m,t} = \mathcal{S}_{t,r}[m]^n$, where $0 \leqslant r \leqslant t$, and $r = 0$ if $m > t+1$.

Proof. Suppose $\mathcal{F} \subset [m]^n$ is compressed and t-intersecting with $\mu(\mathcal{F}) \geqslant (1-\delta)\mu(\mathcal{S}_{n,m,t})$, where $\delta \ll m^{-1}, t^{-1}, \varepsilon$. We consider $\widetilde{\mathcal{F}} \subset \{0,1\}^n$ given by Definition 4.2. By Fact 4.3, $\widetilde{\mathcal{F}}$ is t-intersecting, and $\mu_p(\widetilde{\mathcal{F}}) \geqslant \mu(\mathcal{F}) \geqslant (1-\delta)\mu(\mathcal{S}_{n,m,t}) = (1-\delta)\mu_p(\mathcal{S}_{n,p,t})$, where $p = 1/m$. By Theorem 4.4, we have $\mu_p(\mathcal{F} \setminus \widetilde{\mathcal{S}}) \leqslant \varepsilon\mu(\widetilde{\mathcal{S}})$ for some copy $\widetilde{\mathcal{S}}$ of $\mathcal{S}_{n,p,t} = \mathcal{S}_{t,r}\{0,1\}^n$, where $0 \leqslant r \leqslant t$ (as $p = 1/m \leqslant 1/3$) and $r = 0$ if $m > t+1$ (taking $\zeta < \frac{1}{t+1} - \frac{1}{t+2}$).

We show that the conclusion of the claim holds for $\mathcal{S} = \{x : h^{\otimes n}(x) \in \widetilde{\mathcal{S}}\}$, where $h : [m]^n \to \{0,1\}^n$ is as in Definition 4.2. To see this, first note that $\mathcal{S}$ is a copy of $\mathcal{S}_{n,m,t}$. Furthermore, if $x \in \mathcal{F} \setminus \mathcal{S}$, then $h(x) \in \widetilde{\mathcal{F}} \setminus \widetilde{\mathcal{S}}$, and if x is uniformly random in $[m]^n$, then $h^{\otimes n}(x)$ is distributed as μ_p, so

$$\Pr_{x \in [m]^n}[x \in \mathcal{F} \setminus \mathcal{S}] \leqslant \Pr_{x \in [m]^n}\left[h^{\otimes n}(x) \in \widetilde{\mathcal{F}} \setminus \widetilde{\mathcal{S}}\right] = \mu_p(\widetilde{\mathcal{F}} \setminus \widetilde{\mathcal{S}}) \leqslant \varepsilon\mu_p(\widetilde{\mathcal{S}}) = \varepsilon\mu(\mathcal{S}).$$ □

4.3 | Decompression and local stability

In this subsection, we prove Theorem 2.2 in general, deducing it from the compressed case proved in the previous subsection, and for that, we use decompression and local stability arguments. We

† We state it in a weaker form where we do not specify the exact dependency between parameters, as we do not require this.

start with a proof sketch, where for simplicity, we assume that $m > t+1$, so that the extremal examples are cubes of co-dimension t.

Suppose $\mathcal{F} \subset [m]^n$ is t-intersecting with size close to the maximum possible. Let $\mathcal{G} = T(\mathcal{F})$ be the compressed form of $\mathcal{F}$. By the previous subsection, $\mathcal{G}$ is close to a subcube, say $\mathcal{S} = \{x \mid x_1 = \cdots = x_t = 1\}$.

We now decompress: we consider how the family changes as we undo the compression operators one by one. First, we note that undoing $T_n, T_{n-1}, \dots, T_{t+1}$ does not change the distance from $\mathcal{S}$, so $\mathcal{G}_t = T_1 \circ \dots \circ T_t(\mathcal{F})$ has the same distance to $\mathcal{S}$ as $\mathcal{G}$.

The main point of the argument is to analyse the effect of undoing T_i for $i = 1, \dots, t$. For $j \in [m]$, we let α_j be the fraction of $\mathcal{G}_{t-1}$ with prefix $(1^{t-1}, j)$. If there is some $j^\star$ with $\alpha_{j^\star}$ close to 1, then $\mathcal{G}_{t-1}$ is close to a subcube, and we can continue decompressing. Otherwise, we can partition most of $\mathcal{G}_{t-1}$ into two non-negligible parts such that the value of j in the prefix $(1^{t-1}, j)$ always differs between the two parts. However, as $\mathcal{F}$ is t-intersecting, this implies that the two parts must be cross-intersecting on the coordinates $[n] \setminus [t]$; this will give a contradiction by the following form of Hoffman's bound (which we will prove later in a more general form, see Lemma 5.9).

Lemma 4.6. *Suppose $\mathcal{G}_1, \mathcal{G}_2 \subset [m]^n$ are cross-intersecting with $\mu(\mathcal{G}_i) = \alpha_i$ for $i = 1, 2$. Then $\alpha_1\alpha_2 \leqslant (1-\alpha_1)(1-\alpha_2)/(m-1)^2$.*

We start with a lemma that applies Lemma 4.6 to implement the idea discussed in the previous paragraph. Recall that the largest intersecting codes in $[m]^n$ are the 'dictators' $D_{i\to j} = \{x : x_i = j\}$. We show that if $\mathcal{A}, \mathcal{B} \subset [m]^n$ have nearly maximum size, are cross-intersecting and $T_i(\mathcal{A}), T_i(\mathcal{B}) \subset D_{i\to 1}$, then there is some dictator $D_{i\to j}$ that essentially contains $\mathcal{A}$ and $\mathcal{B}$. Recall that $p = 1/m$ throughout.

Lemma 4.7. *Let $0 < \varepsilon \leqslant 1/15$ and $m \geqslant 3$. Suppose $\mathcal{A}, \mathcal{B} \subset [m]^n$ are cross-intersecting with $\mu(\mathcal{A}), \mu(\mathcal{B}) \geqslant (1-\varepsilon)p$ and $T_i(\mathcal{A}), T_i(\mathcal{B}) \subset D_{i\to 1}$. Then there is $j \in [m]$ such that $\mu(\mathcal{A} \cap D_{i\to j}), \mu(\mathcal{B} \cap D_{i\to j}) \geqslant (1-3\varepsilon)p$.*

Proof. Without loss of generality, we can assume $i = 1$. As $T_1(\mathcal{A}) \subset D_{1\to 1}$, we can write $T_1(\mathcal{A})$ as the disjoint union over $j \in [m]$ of $\mathcal{A}_j := \{(1, z) : z \in \mathcal{A}_{1\to j}\}$; in particular, $\mathcal{A}_1, \dots, \mathcal{A}_m$ are disjoint. Similarly, we may define $\mathcal{B}_1, \dots, \mathcal{B}_m$ and have that $\mathcal{B}_1, \dots, \mathcal{B}_m$ are disjoint. For each $j \in [m]$, let $\alpha_j = \mu(\mathcal{A}_{1\to j})$ and $\beta_j = \mu(\mathcal{B}_{1\to j})$. Then $\sum_j \alpha_j = p^{-1}\mu(\mathcal{A}) \geqslant 1-\varepsilon$ and $\sum_j \beta_j = p^{-1}\mu(\mathcal{B}) \geqslant 1-\varepsilon$. We need to show that for some $j \in [m]$, we have $\alpha_j, \beta_j \geqslant 1-3\varepsilon$.

To see this, suppose without loss of generality that $\alpha = \alpha_1$ is largest among $\{\alpha_j\}_{j\in[m]} \cup \{\beta_j\}_{j\in[m]}$. Let $\mathcal{B}_{\neq 1} := \bigcup_{j\in[2,m]} \mathcal{B}_j$ and $\beta_{\neq 1} = \mu(\mathcal{B}_{\neq 1})$. Then $\beta_{\neq 1} = \sum_{j\neq 1} \beta_j \geqslant 1-\varepsilon-\beta_1 \geqslant 1-\alpha_1-\varepsilon$. As $\mathcal{A}_1$ and $\mathcal{B}_{\neq 1}$ are cross-intersecting, by Lemma 4.6, we have

$$\alpha_1(1-\alpha_1-\varepsilon) \leqslant \alpha_1\beta_{\neq 1} \leqslant (1-\alpha_1)(\beta_1+\varepsilon)/(m-1)^2 \leqslant (1-\alpha_1)(\alpha_1+\varepsilon)/(m-1)^2.$$

Rearranging gives $((m-1)^2-1)\alpha_1(1-\alpha_1) \leqslant ((m-1)^2\alpha_1 + (1-\alpha_1))\varepsilon \leqslant (m-1)^2\varepsilon$. Thus, $\alpha_1(1-\alpha_1) \leqslant 4\varepsilon/3$, so either $\alpha_1 \leqslant 3\varepsilon$ or $\alpha_1 \geqslant 1-3\varepsilon$. We will show that the second bound holds.

Suppose otherwise. Then $\alpha_j \leqslant 3\varepsilon$ for all $j \in [m]$. We can partition $[m]$ as $Q_1 \cup Q_2$ so that for $k = 1, 2$, we have $\sum_{j\in Q_k} \alpha_j \geqslant (1-\varepsilon-3\varepsilon)/2 \geqslant 1/2-2\varepsilon$. Without loss of generality, $\sum_{j\in Q_1} \beta_j \geqslant 1/2-\varepsilon$. Then $\cup_{j\in Q_1} \mathcal{B}_j$ and $\cup_{j\in Q_2} \mathcal{A}_j$ cross-intersect and both have densities at least $1/2-2\varepsilon$ in $[m]^{n-1}$, which contradicts Lemma 4.6, as $\varepsilon \leqslant 1/15$. Thus, $\alpha_1 \geqslant 1-3\varepsilon$, as required. Now we apply

Lemma 4.6 again to $\mathcal{A}_1$ and $\mathcal{B}_{\neq 1}$, which gives $(1-3\varepsilon)\beta_{\neq 1} \leqslant \alpha_1\beta_{\neq 1} \leqslant (1-\alpha_1)(1-\beta_{\neq 1})/(m-1)^2 \leqslant \frac{3\varepsilon}{4}$, so $\beta_{\neq 1} \leqslant 2\varepsilon$. As $\beta_1 \geqslant (1-\varepsilon) - \beta_{\neq 1}$, this gives $\beta_1 \geqslant 1-3\varepsilon$, completing the proof. □

We conclude this subsection with the proof of the stability theorem.

Proof of Theorem 2.2. Let $0 < \delta \ll \varepsilon' \ll \varepsilon, t^{-1}, m^{-1}$. Suppose $\mathcal{F} \subset [m]^n$ is t-intersecting with $\mu(\mathcal{F}) \geqslant (1-\delta)\mu(\mathcal{S}_{n,m,t})$. We can assume without loss of generality that for each $i \in [n]$, the most popular value of x_i for $x \in \mathcal{F}$ is 1 (otherwise we simply relabel the alphabet in that coordinate). We set $\mathcal{F}_0 = F$ and for each $i \in [n]$, let $\mathcal{F}_i = T_i(\mathcal{F}_{i-1})$. By Fact 4.1, $\mathcal{F}_n = T(\mathcal{F}_0)$ is t-intersecting, compressed, and $\mu(\mathcal{F}_n) \geqslant (1-\delta)\mu(\mathcal{S}_{n,m,t})$. By Claim 4.5, $\mu(\mathcal{F} \setminus \mathcal{S}) \leqslant \varepsilon'\mu(\mathcal{S})$ for some copy $\mathcal{S}$ of $\mathcal{S}_{n,m,t} = \mathcal{S}_{t,r}[m]^n$, where $0 \leqslant r \leqslant t$, and $r = 0$ if $m > t+1$. We write J for the set of coordinates on which it depends, so $|J| = t + 2r$.

We define ε_i for all $i \in [n]$ by $\mu(\mathcal{F}_i \cap \mathcal{S}) = (1-\varepsilon_i)\mu(\mathcal{S})$. We note that $\mu(\mathcal{F}) = \mu(\mathcal{F}_n) \leqslant (1+\varepsilon')\mu(\mathcal{S})$ and $\mu(\mathcal{F}_n \cap \mathcal{S}) \geqslant (1-\varepsilon'-\delta)\mu(\mathcal{S}) \geqslant (1-2\varepsilon')\mu(\mathcal{S})$, so $\varepsilon_n \leqslant 2\varepsilon'$. We will show inductively that ε_i is suitably small for $i = n, n-1, \dots, 0$. To prove the theorem, it suffices to show $\varepsilon_0 < \varepsilon/2$, as $\mu(\mathcal{F} \setminus \mathcal{S}) \leqslant \mu(\mathcal{F}) - \mu(\mathcal{F} \cap \mathcal{S}) \leqslant (1+\varepsilon')\mu(\mathcal{S}) - (1-\varepsilon_0)\mu(\mathcal{S}) \leqslant (\varepsilon'+\varepsilon_0)\mu(\mathcal{S})$.

Note that if $i \notin J$, then $\mathcal{S}$ is *i-insensitive*, meaning that for all $x \in [m]^n$, membership of x in $\mathcal{S}$ does not depend on x_i. For such i, we have $|T_i(\mathcal{G}) \cap \mathcal{S}| = |\mathcal{G} \cap \mathcal{S}|$ for any $\mathcal{G} \subset [m]^n$, so $\mu(\mathcal{F}_{i-1} \cap \mathcal{S}) = \mu(\mathcal{F}_i \cap \mathcal{S})$, that is, $\varepsilon_i = \varepsilon_{i-1}$. For $i \in J$, we will show that $\varepsilon_{i-1} \leqslant 3(t+1)^{3t}\varepsilon_i$. This will imply $\varepsilon_0 < (3(t+1)^{3t})^{|J|}\varepsilon' < \varepsilon/2$ as $\varepsilon' \ll \varepsilon$, and so will suffice to complete the proof of the theorem.

Set $J_i := J \setminus \{i\}$. Given $y \in [m]^{J_i}$ and $\mathcal{D} \subset [m]^n$, we use the abbreviation

$$\mathcal{D}(y) := \mathcal{D}_{J_i \to y} = \left\{ z \in [m]^{[n]\setminus J_i} \,\middle|\, (y,z) \in \mathcal{D} \right\} \subset [m]^{[n]\setminus J_i}.$$

We require the following claim, showing that if two J_i-restrictions $\mathcal{F}_i(y_1)$ and $\mathcal{F}_i(y_2)$ are close to the same i-dictator, where y_1, y_2 have agreement at most $t-1$, then this is also true of $\mathcal{F}_{i-1}(y_1)$ and $\mathcal{F}_{i-1}(y_2)$.

Claim 4.8. Suppose $y_1, y_2 \in [m]^{J_i}$ with $\mathsf{agr}(y_1, y_2) \leqslant t-1$ and $\mu(\mathcal{F}_i(y_k) \cap D_{i\to 1}) \geqslant (1-\xi)p$ for both $k = 1, 2$, where $0 \leqslant \xi \leqslant 1/6$. Then there is $j \in [m]$ such that both $\mu(\mathcal{F}_{i-1}(y_k) \cap D_{i\to j}) \geqslant (1-3\xi)p$. Moreover, for any $j' \neq j$, both $\mu(\mathcal{F}_{i-1}(y_k) \cap D_{i\to j'}) < (1-3\xi)p$.

Proof. Note that $\mathcal{F}_{i-1}(y_1)$ and $\mathcal{F}_{i-1}(y_2)$ are cross-intersecting, as $\mathcal{F}_{i-1}$ is t-intersecting and $\mathsf{agr}(y_1, y_2) \leqslant t-1$. Let $\mathcal{A}_k = \{x \in \mathcal{F}_{i-1}(y_k) : T_i(x) \in \mathcal{F}_i(y_k) \cap D_{i\to 1}\}$ for $k = 1, 2$. Then $\mathcal{A}_1, \mathcal{A}_2$ are cross-intersecting and both $\mu(\mathcal{A}_k) > (1-\xi)p$, so the existence of j follows from Lemma 4.7.

For the 'moreover' part, note that if $j' \neq j$, then $(\mathcal{F}_{i-1})_{J\to(y_1,j')}$ and $(\mathcal{F}_{i-1})_{J\to(y_2,j)}$ are cross-intersecting, and applying Lemma 4.6 gives us that $\mu((\mathcal{F}_{i-1})_{J\to(y_1,j')}) < 1-3\xi$. The same argument works interchanging the roles of y_1 and y_2, and we get that $\mu(\mathcal{F}_{i-1}(y_k) \cap D_{i\to j'}) < (1-3\xi)p$. □

Using Claim 4.8, we now bound ε_{i-1}. We start with the case $r = 0$. Let $\mathbf{1} \in [m]^{J_i}$ be the all-1 vector. We have $\mu(\mathcal{F}_i(\mathbf{1}) \cap D_{i\to 1}) = p^{1-t}\mu(\mathcal{F}_i \cap \mathcal{S}) = (1-\varepsilon_i)p$, so by Claim 4.8 with $y_1 = y_2 = \mathbf{1}$, there is $j \in [m]$ such that $\mu(\mathcal{F}_{i-1}(\mathbf{1}) \cap D_{i\to j}) \geqslant (1-3\varepsilon_i)p$. The most popular value in $\mathcal{F}_{i-1}$ of coordinate i is 1 (since this is the case in $\mathcal{F}$), so $j = 1$. We deduce $\mu(\mathcal{F}_{i-1} \cap \mathcal{S}) = p^{t-1}\mu(\mathcal{F}_{i-1}(\mathbf{1}) \cap D_{i\to 1}) \geqslant (1-3\varepsilon_i)p^t$, so $\varepsilon_{i-1} \leqslant 3\varepsilon_i$.

It remains to consider $r \geqslant 1$. We have $m \leqslant t+1$ by Claim 4.5. For a vector $y \in [m]^n$ and $j \in [m]$, let $y[j]$ be the set of coordinates of i equal to j. We partition $\mathcal{S}$ as $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1$, where

$$\mathcal{S}_0 := \{x \in [m]^n : x|_{J_i} \in \mathcal{G}_0\}, \qquad \mathcal{G}_0 := \{y \in [m]^{J_i} : |y[1]| > t+r-1\},$$
$$\mathcal{S}_1 = \{x \in [m]^n : x|_{J_i} \in \mathcal{G}_1\} \cap D_{i\to 1}, \qquad \mathcal{G}_1 := \{y \in [m]^{J_i} : |y[1]| = t+r-1\}.$$

As $\mathcal{S}_0$ is i-insensitive, $\mu(\mathcal{F}_{i-1} \cap \mathcal{S}_0) = \mu(\mathcal{F}_i \cap \mathcal{S}_0)$. Now we wish to show that $\mu(\mathcal{F}_{i-1} \cap \mathcal{S}_1)$ is large, that is, that $\mu(\mathcal{F}_{i-1}(y) \cap D_{i\to 1})$ is close to 1 for each $y \in \mathcal{G}_1$. Firstly we show this for $\mathcal{F}_i$.

Claim 4.9. $\mu(\mathcal{F}_i(y) \cap D_{i\to 1}) \geqslant (1-(t+1)^{3t}\varepsilon_i)p$ for each $y \in \mathcal{G}_1$.

Proof. To see this, we note that

$$\varepsilon_i \mu(\mathcal{S}) = \mu(\mathcal{S} \setminus \mathcal{F}_i) \geqslant \mu(\mathcal{S}_1 \setminus \mathcal{F}_i) = \sum_{y\in\mathcal{G}_1} p^{|J_i|}(p - \mu(\mathcal{F}_i(y) \cap D_{i\to 1})).$$

Each summand on the right-hand side is non-negative, and $|J| = t+2r \leqslant 3t$, so for each $y \in \mathcal{G}_1$, we have $p - \mu(\mathcal{F}_i(y) \cap D_{i\to 1})) \leqslant p(t+1)^{3t}\varepsilon_i$, so the claim holds. $\square$

Next, we prove the corresponding claim for $\mathcal{F}_{i-1}$, although at first just with $D_{i\to j}$ for some $j \in [m]$; the theorem will follow once we show that $j=1$. We say that $y \in \mathcal{G}_1$ is j-good if $\mu(\mathcal{F}_{i-1}(y) \cap D_{i\to j}) \geqslant (1-3(t+1)^{3t}\varepsilon_i)p$.

Claim 4.10. There is some $j \in [m]$ such that every $y \in \mathcal{G}_1$ is j-good.

Proof. Note that for any $y, y' \in \mathcal{G}_1$ with $\mathrm{agr}(y,y') = t-1$, by Claims 4.8 and 4.9, there is some $j \in [m]$ such that both y and y' are j-good. The claim then follows from the observation that the graph G whose edges consist of such pairs $\{y, y'\}$ is connected. (We can get between any two elements of $\mathcal{G}_1$ by a sequence of steps where in each step, we change some coordinate from 1 to another value and some coordinate from another value to 1, and each such step can be implemented by a path of length two in G.) $\square$

It remains to show that $j=1$. We consider $\mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1$, where $\mathcal{T}_0 = \{y \in [m]^n \,|y_{J_i} \in \mathcal{G}_0\} = \mathcal{S}_0$ and $\mathcal{T}_1 = \{y \in [m]^n \,|y_{J_i} \in \mathcal{G}_1, y_i = j\}$. Recalling that $\mu(\mathcal{F}_{i-1} \cap \mathcal{S}_0) = \mu(\mathcal{F}_i \cap \mathcal{S}_0)$, by the previous claim, we deduce $\mu(\mathcal{F}_{i-1} \cap \mathcal{T}) \geqslant (1-3(t+1)^{3t}\varepsilon_i)\mu(\mathcal{T})$, so

$$\mu(\mathcal{F}_{i-1} \setminus \mathcal{T}) \leqslant \mu(\mathcal{F}_{i-1}) - \mu(\mathcal{F}_{i-1} \cap \mathcal{T}) \leqslant 3(t+1)^{3t}\varepsilon_i + \varepsilon' \leqslant 4(t+1)^{3t}\varepsilon_i,$$

where in the second inequality, we used $\mu(\mathcal{F}_{i-1}) = \mu(\mathcal{F}) \leqslant (1+\varepsilon')\mu(\mathcal{S}) = (1+\varepsilon')\mu(\mathcal{T})$. Hence, for $\ell \neq j$, the fraction of $x \in \mathcal{F}_{i-1}$ such that $x_i = \ell$ is at most $4(t+1)^{3t}\varepsilon_i + q(\ell)$, where $q(\ell)$ is the fraction of $x \in \mathcal{T}$ that have $x_i = \ell$; by symmetry, this value is the same for all $\ell \neq j$, and we denote it by q. The fraction of $x \in \mathcal{F}_{i-1}$ such that $x_i = j$ is, for the same reasons, at least $q(j) - 4(t+1)^{3t}\varepsilon_i$. But $q(j) \geqslant q + \mu(\mathcal{G}_1) \geqslant q + (t+1)^{-3t}$, so $q(j) - 4(t+1)^{3t}\varepsilon_i \geqslant q + 4(t+1)^{3t}\varepsilon_i$ (we can ensure $\varepsilon_i < 1/8$). Thus, j is the most popular value of coordinate i in $\mathcal{F}_{i-1}$, so $j=1$, as required. $\square$

4.4 | Bootstrapping

We conclude this part by proving Theorem 2.3, which completes the proof of Theorem 2.4. We will use compressions to reduce to the cube, so we start with some remarks in this setting.

We consider $\{0,1\}^n$ equipped with the uniform measure μ. Suppose $\mathcal{A}, \mathcal{B} \subset \{0,1\}^n$. We say that $\mathcal{A}, \mathcal{B}$ are cross-intersecting if for any $x \in \mathcal{A}$, $y \in \mathcal{B}$, there is $i \in [n]$ such that $x_i = y_i = 1$. We say that $\mathcal{A}, \mathcal{B}$ are cross-agreeing if for any $x \in \mathcal{A}$, $y \in \mathcal{B}$, there is $i \in [n]$ such that $x_i = y_i$. Clearly, if $\mathcal{A}, \mathcal{B}$ are cross-intersecting, then they are cross-agreeing. We have the following easy fact, which is immediate from the observation that if $\mathcal{A}, \mathcal{B}$ are cross-agreeing and $x + y = \mathbf{1}$ (the all-1 vector), then we cannot have $x \in \mathcal{A}$ and $y \in \mathcal{B}$.

Fact 4.11. *If $\mathcal{A}, \mathcal{B} \subset \{0,1\}^n$ are cross-agreeing, then $\mu(\mathcal{A}) + \mu(\mathcal{B}) \leqslant 1$.*

We also require the following isoperimetric lemma of Ellis, Keller and Lifshitz [7].

Lemma 4.12. *Suppose $0 \leqslant p \leqslant q \leqslant 1$, $\alpha \geqslant 0$ and $\mathcal{F} \subset \{0,1\}^n$ is monotone. If $\mu_p(\mathcal{F}) \geqslant p^\alpha$, then $\mu_q(\mathcal{F}) \geqslant q^\alpha$.*

Proof of Theorem 2.3. Let $\mathcal{G}, \mathcal{H} \subset [m]^n$ with $\mu(\mathcal{H}) = m^{-t}\varepsilon$ and $\mu(\mathcal{G}) > 1 - C\varepsilon$, where $0 \ll \varepsilon \ll t^{-1}, C^{-1}$. Suppose for contradiction that $\mathcal{G}$ and $\mathcal{H}$ are cross-agreeing. Let $\mathcal{G}' = \widetilde{T(\mathcal{G})}$ and $\mathcal{H}' = \widetilde{T(\mathcal{H})}$, where T is the compression operator and the operator $\mathcal{F} \to \widetilde{\mathcal{F}}$ is from Definition 4.2. By Facts 4.1 and 4.3, $\mathcal{G}'$ and $\mathcal{H}'$ are monotone and cross-intersecting with $\mu_p(\mathcal{G}') \geqslant \mu(\mathcal{G})$ and $\mu_p(\mathcal{H}') \geqslant \mu(\mathcal{H})$, where $p = 1/m$.

Now we consider $\mathcal{G}'$ and $\mathcal{H}'$ under the uniform measure $\mu = \mu_{1/2}$. By monotonicity, we have $\mu(\mathcal{G}') \geqslant \mu_p(\mathcal{G}') > 1 - C\varepsilon$. By Lemma 4.12, $\mu(\mathcal{H}') \geqslant \mu_p(\mathcal{H}')^{\log_p(1/2)}$, so $\log_2 \mu(\mathcal{H}') \geqslant \log_2(m^{-t}\varepsilon)/\log_2(m)$, giving $\mu(\mathcal{H}') \geqslant 2^{-t}\varepsilon^{0.7}$, as $m \geqslant 3$ and $1/\log_2(3) < 0.7$. However, $1 - C\varepsilon + 2^{-t}\varepsilon^{0.7} > 1$ as $\varepsilon \ll t^{-1}, C^{-1}$, which contradicts Fact 4.11. □

Part II: Moderate alphabets

In this part, we prove our main result Theorem 1.1 when the alphabet size m is moderate, that is, $m_0(t) \leqslant m \leqslant 2^{n/N(t)}$. Here the largest t-intersecting codes are subcubes of codimension t. As mentioned in the introduction, we cannot achieve such a strong pseudorandomness condition in our regularity lemma as in the case of fixed m, so we settle for a weaker notion of 'uncapturability'. We present the proof in the second section of the part, where to implement a version of the cross-agreement strategy, we introduce a gluing argument that exploits expansion under another pseudorandomness condition, namely globalness. The tools for this are developed in the first section, in which we study our two pseudorandomness conditions (uncapturability and globalness) and establish the small-set expansion for global functions via a refined version of our global hypercontractivity inequality from [16].

5 | TOOLS

This section concerns various properties of the pseudorandomness notions of uncapturability and globalness, particularly a regularity lemma for uncapturability and a small set expansion property for global functions, which is analogous to Theorem 3.3. The latter will be established via a corresponding statement for the noise operator, which will be proved by a refined form of our global hypercontractivity inequality. Along the way, we record various facts needed here and later concerning Markov chains and the Efron–Stein theory of orthogonal decompositions.

5.1 | Uncapturability and globalness

This subsection contains the definitions and basic properties of the two key pseudorandomness conditions used in this part. We start with uncapturability, which is the condition that will appear in the regularity lemma in the next subsection. Recall that for $\mathcal{F} \subset [m]^n$ and $\alpha \in [m]^R$ for some $R \subset [n]$, we write $\mathcal{F}[\alpha] = \{x \in \mathcal{F} : x_R = \alpha\}$ and $\mathcal{F}_{R\to\alpha} = \mathcal{F}(\alpha) = \{x \in [m]^{[n]\setminus R} : (x,\alpha) \in \mathcal{F}\}$. We also write $D_{R\to\alpha} = \{x \in [m]^n : x_R = \alpha\}$, which is a subcube of co-dimension $|R|$, which we refer to as a 'dictator' if $|R| = 1$. For a collection of subcubes $\mathcal{D}$, we denote by $\bigcup \mathcal{D}$ the union of these subcubes, that is, $\bigcup_{D\in\mathcal{D}} D$.

Definition 5.1. We say $\mathcal{F} \subset [m]^n$ is (r,ε)-capturable if there is a set $\mathcal{D}$ of at most r dictators with $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant \varepsilon$. Otherwise, we say that $\mathcal{F}$ is (r,ε)-uncapturable.

Now we define the stronger (see Claim 5.4) condition of globalness.

Definition 5.2. We say $f : [m]^n \to \mathbb{R}$ is (r,ε)-global if for any $R \subset [n]$ with $|R| \leqslant r$ and $a \in [m]^R$, we have $\|f_{R\to a}\|_2^2 \leqslant \varepsilon$. We say $\mathcal{F} \subset [m]^n$ is (r,ε)-global if its characteristic function is (r,ε)-global.

Most of this section will be devoted to the proof of the following small set expansion property for global functions, which is analogous to Theorem 3.3. We remark that we will later use Theorem 5.3 to prove that random gluings significantly increase the measure of global families.

Theorem 5.3. *For any $\lambda > 0$, there is $c > 0$ such that the following holds for Markov chains T_i on Ω_i with $\lambda_*(T_i) \geqslant \lambda$ for all $i \in [n]$ and consecutive random states x, y of the stationary chain for the product chain T on Ω. If $\mathcal{F} \subset \Omega$ is $(\log(1/\mu), \mu^{1-c})$-global with $\mu \in (0, 1/16)$, then $\mathbb{P}(x \in \mathcal{F}, y \in \mathcal{F}) \leqslant \mu^{1+c}$.*

We begin by giving two simple relations between uncapturability and globalness that will be useful for us. The first property asserts that globalness implies very strong uncapturability.

Claim 5.4. If $\gamma \in (0,1)$ and $\emptyset \neq \mathcal{G} \subset [m]^n$ is $(1, \mu(\mathcal{G})/\gamma)$-global, then $\mathcal{G}$ is $(\gamma m/4, \mu(\mathcal{G})/2)$-uncapturable.

Proof. Suppose that $\mathcal{D}$ is a set of dictators with $\mu(\mathcal{G} \setminus \bigcup \mathcal{D}) \leqslant \mu(\mathcal{G})/2$. We need to show $|\mathcal{D}| > \gamma m/4$. By assumption $\mu(\mathcal{G}_{i\to a}) \leqslant \mu(\mathcal{G})/\gamma$ for each $D_{i\to a} \in \mathcal{D}$, so by a union bound

$$\mu(\mathcal{G})/2 \leqslant \mu(\mathcal{G} \cap \bigcup \mathcal{D}) \leqslant \sum_{D_{i\to a}\in\mathcal{D}} \mu(\mathcal{G} \cap D_{i\to a}) = \sum_{D_{i\to a}\in\mathcal{D}} m^{-1} \cdot \mu(\mathcal{G}_{i\to a}) \leqslant |\mathcal{D}| m^{-1} \frac{\mu(G)}{\gamma}.$$

Thus, $|\mathcal{D}| \geqslant \gamma m/2 > \gamma m/4$, as required. $\square$

The second property shows that any family $\mathcal{G}$ with significant measure can be made global by taking small restrictions.

Lemma 5.5. *Let $0 < \gamma < 1$ and $r, m, n \in \mathbb{N}$. For any $\mathcal{G} \subset [m]^n$, there is $R \subset [n]$ and $\alpha \in [m]^R$ with $|R| \leqslant r \log_{\gamma^{-1}}(\mu(\mathcal{G})^{-1})$ such that $\mathcal{G}' = \mathcal{G}_{R\to\alpha}$ is $(r, \mu(\mathcal{G}')/\gamma)$-global with $\mu(\mathcal{G}') \geqslant \mu(\mathcal{G})$.*

Proof. Starting with $\mathcal{G}_0 = \mathcal{G}$, for each $i \geqslant 0$, if $\mathcal{G}_i$ is not $(r, \mu(\mathcal{G}_i)/\gamma)$-global, we let $\mathcal{G}_{i+1} = (\mathcal{G}_i)_{R_i\to\alpha_i}$ for some $R_i \subset [n]$ with $|R_i| \leqslant r$ and $\alpha_i \in [m]^{R_i}$ such that $\mu(\mathcal{G}_{i+1}) \geqslant \mu(\mathcal{G}_i)/\gamma$; such a restriction exists by definition. As all measures are bounded by 1, there can be at most $\log_{\gamma^{-1}}(\mu(\mathcal{G})^{-1})$ iterations, at which point we terminate with $\mathcal{G}' = \mathcal{G}_{R\to\alpha}$ with the stated properties. □

5.2 | The uncapturable code regularity lemma

In this subsection, we prove the following regularity lemma which approximately decomposes any code into pieces corresponding to uncapturable restrictions.

Lemma 5.6. *Let $r, k, m \in \mathbb{N}$ and $\varepsilon \geqslant 1/m$. For any $\mathcal{F} \subset [m]^n$, there is a collection $\mathcal{D}$ of at most r^k subcubes of co-dimension at most k such that $\mathcal{F}_{R\to\alpha}$ is $(r, \varepsilon\mu(D)^{-1}m^{-k})$-uncapturable for each $D = D_{R\to\alpha} \in \mathcal{D}$ and $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant 3r^{k+1}\varepsilon m^{-k}$.*

Proof. We may assume that $\mathcal{F}$ is $(r, \varepsilon m^{-k})$-capturable; otherwise, the lemma holds with $\mathcal{D} = \{[m^n]\}$. We apply the following iterative process for $s = 1, \dots, k$.

- We let $\mathcal{D}'_{s-1}$ be the set of $D = D_{R\to\alpha} \in \mathcal{D}_{s-1}$ such that $\mathcal{F}_{R\to\alpha}$ is $(r, \varepsilon\mu(D)^{-1}m^{-k})$-capturable, where for $s = 1$, we let $\mathcal{D}'_0 = \mathcal{D}_0 = \{D_{\emptyset\to\emptyset}\} = \{[m]^n\}$.
- For each $D = D_{R\to\alpha} \in \mathcal{D}'_{s-1}$, by definition of capturability, we can fix a set $\mathcal{D}[D]$ of at most r dictators such that $\mu(\mathcal{F}_{R\to\alpha} \setminus \bigcup \mathcal{D}[D]) \leqslant \varepsilon\mu(D)^{-1}m^{-k}$.
- We define $\mathcal{D}_s = \{D_{(R,i)\to(\alpha,a)} : D = D_{R\to\alpha} \in \mathcal{D}'_{s-1}, D_{i\to a} \in \mathcal{D}[D]\}$.

At the end of the process, we let $\mathcal{D}'_k \subset \mathcal{D}_k$ be the set of $D = D_{R\to\alpha} \in \mathcal{D}_k$ such that $\mathcal{F}_{R\to\alpha}$ is (r, ε)-capturable. We will show that $\mathcal{D} = (\mathcal{D}_1 \setminus \mathcal{D}'_1) \cup \dots \cup (\mathcal{D}_k \setminus \mathcal{D}'_k)$ satisfies the requirements of the lemma.

Clearly, for all $D \in \mathcal{D}$, we have that $\mathcal{F} \cap D$ is $(r, \varepsilon\mu(D)^{-1}m^{-k})$-uncapturable, and $|\mathcal{D}| \leqslant r^k$ as we explore at most many subcubes during the above process. We will bound $\mu(\mathcal{F} \setminus \bigcup \mathcal{D})$ by $\mu(\mathcal{F} \setminus \bigcup(\mathcal{D} \cup \mathcal{D}'_k)) + \mu(\mathcal{F} \cap \bigcup \mathcal{D}'_k)$.

For the first term in the bound, we write $\mathcal{F} \setminus \bigcup(\mathcal{D} \cup \mathcal{D}'_k) = \cup_{s=0}^{k-1} \mathcal{E}_s$, where each

$$\mathcal{E}_s = \bigcup\{\mathcal{F}_{R\to\alpha} \setminus \bigcup \mathcal{D}[D] : D = D_{R\to\alpha} \in \mathcal{D}'_s\}.$$

By definition, $\mu(\mathcal{E}_s) \leqslant \sum_{D\in\mathcal{D}'_s} \mu(D) \cdot \varepsilon\mu(D)^{-1}m^{-k} = |\mathcal{D}'_s|\varepsilon m^{-k}$, so

$$\mu(\mathcal{F} \setminus \bigcup(\mathcal{D} \cup \mathcal{D}'_k)) \leqslant \varepsilon m^{-k} \sum_{s=0}^{k-1} |\mathcal{D}'_s| \leqslant r^k \varepsilon m^{-k}.$$

For the second term in the bound, we note that if $D_{R\to\alpha} \in \mathcal{D}'_k$, then $\mathcal{F}_{R\to\alpha}$ is (r,ε)-capturable, so has measure is at most $r\frac{1}{m} + \varepsilon \leqslant (r+1)\varepsilon$. Thus,

$$\mu(\mathcal{F} \cap \bigcup \mathcal{D}'_k) \leqslant \sum_{D\in\mathcal{D}'_k} \mu(D)(r+1)\varepsilon \leqslant r^k m^{-k}(r+1)\varepsilon.$$

We deduce $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant \mu(\mathcal{F} \setminus \bigcup(\mathcal{D} \cup \mathcal{D}'_k)) + \mu(\mathcal{F} \cap \bigcup \mathcal{D}'_k) \leqslant 3r^{k+1}\varepsilon m^{-k}$. □

5.3 | Markov chains and orthogonal decompositions

This subsection contains some further theory of Markov chains, Efron–Stein orthogonal decompositions and a general form of the Hoffman bound for cross-intersecting families in any product space. The results are somewhat standard, but we include details for the convenience of the reader.

Let T be a Markov chain on S with stationary distribution ν. The absolute spectral gap $\lambda_* = \lambda_*(T)$ is

$$(1-\lambda_*)^2 = \sup\{\mathbb{E}(Tf)^2 : \mathbb{E}f = 0, \mathbb{E}f^2 = 1\}.$$

Here expectations are with respect to ν. If T is reversible, we can also view λ_* as the minimum value of $1 - |\lambda|$ over all eigenvalues $\lambda \neq 1$. We start with a general lower bound for λ_*.

Lemma 5.7. *Let T be a Markov chain on S with stationary distribution ν such that $T_{ab} \geqslant \alpha\nu(b)$ for every $a, b \in S$. Then $\lambda_*(T) \geqslant \alpha$.*

Proof. By assumption, $S_{ab} := T_{ab} - \alpha\nu(b) \geqslant 0$, with $\sum_b S_{ab} = 1 - \alpha$ and $\sum_a \nu(a)S_{ab} = (1-\alpha)\nu(b)$.

If $\mathbb{E}f = 0$ and $\mathbb{E}f^2 = 1$, then by Cauchy–Schwarz

$$\begin{aligned}
\mathbb{E}(Tf)^2 &= \sum_a \nu(a)(Tf)(a)^2 = \sum_a \nu(a)(\sum_b T_{ab}f(b))^2 = \sum_a \nu(a)\left(\sum_b S_{ab}f(b)\right)^2 \\
&\leqslant \sum_a \nu(a)\left(\sum_b S_{ab}\right)\left(\sum_b S_{ab}f(b)^2\right) = (1-\alpha)\sum_a \nu(a)\sum_b S_{ab}f(b)^2 \\
&= (1-\alpha)\sum_b f(b)^2 \sum_a \nu(a)S_{ab} = (1-\alpha)^2 \sum_b \nu(b)f(b)^2 = (1-\alpha)^2.
\end{aligned}$$

□

Now we consider Markov chains T_i acting on Ω_i for $i \in [n]$ and their tensor product $T = T_1 \otimes \cdots \otimes T_n$ acting on $\Omega = \Omega_1 \otimes \cdots \otimes \Omega_n$, with transition matrix $T_{xy} = \prod_{i=1}^n (T_i)_{x_i y_i}$. The stationary distribution of T is $\nu = \nu_1 \otimes \cdots \otimes \nu_n$, where each ν_i is stationary for T_i. We will often have $\Omega = [m]^n$ and ν uniform, but we will also require the general setting.

We use the Efron–Stein orthogonal decomposition (see, e.g. [27, Section 8.3]): for any $f \in L^2(\Omega, \nu)$, we can write $f = \sum_{S\subset[n]} f^{=S}$, where each $f^{=S}$ is characterised by the properties that it only depends on coordinates in S and that it is orthogonal to any function which depends only on some set of coordinates not containing S; in particular, $f^{=S}$ and $f^{=S'}$ are orthogonal for $S \neq S'$.

We have similar Plancherel / Parseval relations as for Fourier decompositions, namely $\langle f, g\rangle = \sum_S \mathbb{E}[f^{=S}g^{=S}]$, so $\mathbb{E}[f^2] = \sum_S \mathbb{E}[(f^{=S})^2]$. Explicitly, we let $f^{\subset J}(x) = \mathbb{E}_{y\sim\nu}[f(y) \mid y_{\overline{J}} = x_{\overline{J}}]$, and then, we have $f^{=S} = \sum_{J\subset S}(-1)^{|S\setminus J|} f^{\subset J}$ (the inclusion–exclusion formula for $f^{\subset J} = \sum_{S\subset J} f^{=S}$). We note the following identity which is immediate from this construction.

Fact 5.8. *For $S \subset T \subset [n]$, $x \in \Omega_S$ and $f \in L^2(\Omega, \nu)$, we have $(f^{=T})_{S\to x} = (f_{S\to x})^{=T\setminus S}$.*

We require the following general form of the well-known Hoffman bound (the uniform case was used in Part I, see Lemma 4.6). We include the proof for completeness.

Lemma 5.9. *Let $\nu = \prod_{i=1}^n \nu_i$ be a product probability measure on $[m]^n$ such that $\nu_i(x) \leqslant \lambda \leqslant 1/2$ for all $i \in [n]$, $x \in [m]$. Suppose $\mathcal{G}_1, \mathcal{G}_2 \subset [m]^n$ are cross-intersecting with $\nu(\mathcal{G}_i) = \alpha_i$ for $i = 1, 2$. Then*

$$\alpha_1\alpha_2 \leqslant \left(\frac{\lambda}{1-\lambda}\right)^2 (1-\alpha_1)(1-\alpha_2).$$

The proof of Lemma 5.9 requires the following estimate.

Claim 5.10. Let U_i be Markov chains on Ω_i for $i \in [n]$ and let U be the product chain on Ω. For any $f : \Omega \to \mathbb{R}$ and $S \subset [n]$, we have $\|Uf^{=S}\|_2 \leqslant \|f^{=S}\|_2 \prod_{i\in S}(1 - \lambda_*(U_i))$.

Proof. Since $f^{=S}$ does not depend on variables outside S, we may assume without loss of generality that $S = [n]$. We introduce interpolating operators $U_{\leqslant j} = \bigotimes_{i=1}^{j} U_i \otimes \bigotimes_{i=j+1}^{n} I_i$, where I_i is the identity, and $g_j = U_{\leqslant j} f^{=S}$ for $0 \leqslant j \leqslant n$. It suffices to show $\|g_j\|_2 \leqslant (1 - \lambda_*(U_j))\|g_{j-1}\|_2$ for $j \in [n]$.

We calculate $\|g_j\|_2^2 = \mathbb{E}(U_j g_{j-1})^2$ by conditioning on $z \in \Omega_{[n]\setminus\{j\}}$, that is,

$$\mathbb{E}(U_j g_{j-1})^2 = \mathop{\mathbb{E}}_{\mathbf{z}\sim\nu_{[n]\setminus\{j\}}} \left[\mathbb{E}(U_j h_{\mathbf{z}})^2\right],$$

where $h_z := (g_{j-1})_{[n]\setminus\{j\}\to z} \in L^2(\Omega_j, \nu_j)$. Note that for each z,

$$\mathop{\mathbb{E}}_{x\sim\nu_j}[h_z(x)] = \mathop{\mathbb{E}}_{x\sim\nu_j}\left[(U_{\leqslant j-1} f^{=S})(z, x)\right] = U_{\leqslant j-1}\left(\mathop{\mathbb{E}}_{x\sim\nu_j}\left[f^{=S}(z, x)\right]\right) = U_{\leqslant j-1} 0 = 0,$$

so $\mathbb{E}(U_j h_z)^2 \leqslant (1 - \lambda_*(U_j))^2 \mathbb{E}h_z^2$. As $\mathbb{E}_{\mathbf{z}}\mathbb{E}h_{\mathbf{z}}^2 = \mathbb{E}g_{j-1}^2$, we get $\|g_j\|_2 \leqslant (1 - \lambda_*(U_j))\|g_{j-1}\|_2$, as required. □

Proof of Lemma 5.9. For each $i \in [n]$, we consider the Markov chain U_i on $[m]$ with transition probabilities $(U_i)_{xx} = 0$ and $(U_i)_{xy} = \nu_i(y)/(1 - \nu_i(x))$ for $y \neq x$. We claim that

$$1 - \lambda_*(U_i) \leqslant \frac{\lambda}{1-\lambda}. \tag{1}$$

This holds as for any $f \in L^2([m], \nu_i)$ with $\mathbb{E} f = 0$ and $\mathbb{E} f^2 = 1$, we have

$$\|U_i f\|_2^2 = \sum_x \nu_i(x) \left(\sum_{y \neq x} \frac{\nu_i(y)}{1 - \nu_i(x)} f(y) \right)^2 = \sum_x \frac{\nu_i(x)}{(1 - \nu_i(x))^2} \left(\sum_{y \neq x} \nu_i(y) f(y) \right)^2$$
$$= \sum_x \frac{\nu_i(x)}{(1 - \nu_i(x))^2} (\nu_i(x) f(x))^2 \leqslant \left(\frac{\lambda}{1-\lambda} \right)^2 \sum_x \nu_i(x) f(x)^2 = \left(\frac{\lambda}{1-\lambda} \right)^2.$$

Next we note that if $\mathbf{x} \sim \nu$ and $\mathbf{y} \sim U\mathbf{x}$, then $\mathbb{P}(\mathbf{x} \in \mathcal{G}_1, \mathbf{y} \in \mathcal{G}_2) = 0$, as $\mathsf{agr}(\mathbf{x}, \mathbf{y}) = 0$ by definition of U, but $\mathcal{G}_1, \mathcal{G}_2$ are cross-intersecting by assumption. We can also write this probability as $\langle g_1, U g_2 \rangle$, where $g_1, g_2 : [m]^n \to \{0, 1\}$ are the indicator functions of $\mathcal{G}_1, \mathcal{G}_2$. By orthogonality and Cauchy–Schwarz,

$$0 = \sum_{S \subset [n]} \langle g_1^{=S}, U g_2^{=S} \rangle = \alpha_1 \alpha_2 + \sum_{S \neq \emptyset} \langle g_1^{=S}, U g_2^{=S} \rangle \geqslant \alpha_1 \alpha_2 - \sum_{S \neq \emptyset} \left\| g_1^{=S} \right\|_2 \left\| U g_2^{=S} \right\|_2.$$

By Claim 5.10 and (1), we have $\|U g_2^{=S}\| \leqslant (\frac{\lambda}{1-\lambda})^{|S|} \|g_2^{=S}\|$, so

$$\alpha_1 \alpha_2 \leqslant \sum_{S \neq \emptyset} \left(\frac{\lambda}{1-\lambda} \right)^{|S|} \left\| g_1^{=S} \right\|_2 \left\| g_2^{=S} \right\|_2 \leqslant \frac{\lambda}{1-\lambda} \sum_{S \neq \emptyset} \left\| g_1^{=S} \right\|_2 \left\| g_2^{=S} \right\|_2.$$

By Cauchy–Schwarz and Parseval

$$\left(\sum_{S \neq \emptyset} \left\| g_1^{=S} \right\|_2 \left\| g_2^{=S} \right\|_2 \right)^2 \leqslant \sum_{S \neq \emptyset} \left\| g_1^{=S} \right\|_2^2 \sum_{S \neq \emptyset} \left\| g_2^{=S} \right\|_2^2 = \mathrm{Var}(g_1) \mathrm{Var}(g_2) = \alpha_1 (1 - \alpha_1) \alpha_2 (1 - \alpha_2).$$

We deduce $(\alpha_1 \alpha_2)^2 \leqslant (\frac{\lambda}{1-\lambda})^2 \alpha_1 \alpha_2 (1 - \alpha_1)(1 - \alpha_2)$, as required. □

5.4 | Small set expansion via noise stability

The goal for the remainder of this section is to prove Theorem 5.3 concerning global small set expansion. We start by reducing it to the case of a particular Markov chain, namely that given by the noise operator, which we will now define. Let $\nu = \prod_{i=1}^n \nu_i$ be a product probability measure on $\Omega = \prod_{i=1}^n \Omega_i$. Fix $\rho \in [0, 1]$. We let T_i be the Markov chain on Ω_i with transition probabilities $(T_i)_{xy} = \rho 1_{y=x} + (1 - \rho) \nu_i(y)$, that is, from any state x, we stay at x with probability ρ or otherwise move to a random state according to ν_i. We let T be the product chain on Ω. We also write $\mathrm{T} = \mathrm{T}_\rho$. We call T_ρ the *noise operator* when we think of it as an operator on $L^2(\Omega, \nu)$ via $(\mathrm{T}_\rho f)(x) = \mathbb{E}_{\mathbf{y} \sim \mathrm{T}_\rho x}[f(\mathbf{y})]$.

Recall that in Theorem 5.3, we want to bound $\mathbb{P}(x \in \mathcal{F}, y \in \mathcal{F})$ for some $\mathcal{F} \subset \Omega$, when x and y are consecutive states of the stationary chain for some product chain U on Ω. The analytic form is to bound $\langle f, U f \rangle$ where f is the characteristic function of $\mathcal{F}$. We will soon see that this can be bounded by an analogous expression in terms of the noise operator, that is, $\mathsf{Stab}_\rho(f) := \langle f, \mathrm{T}_\rho f \rangle$, which is called the *noise stability* of f. For future reference, we note the following estimate, showing that a bound on the noise stability for any given $\rho > 0$ implies one for all $\rho < 1$.

Lemma 5.11. $\mathrm{Stab}_\rho(f) \leqslant \|f\|_2^{2(1-1/t)} \mathrm{Stab}_{\rho^t}(f)^{1/t}$ *whenever* $t = 2^d$ *with* $d \in \mathbb{N}$.

Proof. By Cauchy–Schwarz, we have

$$\mathrm{Stab}_\rho(f) = \langle f, T_\rho f\rangle \leqslant \|f\|_2 \left\|T_\rho f\right\|_2 = \|f\|_2 \sqrt{\mathrm{Stab}_{\rho^2}(f)}.$$

The lemma follows by iterating this estimate. □

We also need the following well-known formulae for the noise operator and stability.

Fact 5.12. $\mathrm{T}_\rho f(x) = \sum_{S\subset[n]} \rho^{|S|} f^{=S}(x)$ *and* $\mathrm{Stab}_\rho(f) = \sum_{S\subset[n]} \rho^{|S|}\|f^{=S}\|_2^2$.

The following lemma reduces showing small set expansion of a general chain U to that of the noise operator, provided that we have a uniform lower bound on the absolute spectral gap in each coordinate.

Lemma 5.13. *Let* $U = \prod_{i=1}^n U_i$ *be a product chain on* $\Omega = \prod_{i=1}^n \Omega_i$ *with each* $\lambda_*(U_i) \geqslant \lambda$. *Then for all* $f : \Omega \to \mathbb{R}$, *we have* $\langle f, Uf\rangle \leqslant \mathrm{Stab}_{1-\lambda}(f)$.

Proof. We use the orthogonal decomposition $f = \sum_{S\subset[n]} f^{=S}$. We note that $\langle f^{=S}, Uf^{=T}\rangle$ can only be non-zero if $S = T$, as $Uf^{=T}$ only depends on coordinates in T. Thus,

$$\langle f, Uf\rangle = \sum_{S\subset[n]} \langle f^{=S}, Uf^{=S}\rangle \leqslant \sum_{S\subset[n]} \left\|f^{=S}\right\|_2 \left\|Uf^{=S}\right\|_2.$$

Applying Claim 5.10 and Fact 5.12 completes the proof. □

By Lemma 5.13, to prove Theorem 5.3, it remains to prove the following corresponding global small set expansion theorem for the noise operator.

Theorem 5.14. *For every* $\rho < 1$, *there is* $c > 0$ *such that if* $f : \Omega \to \{0,1\}$ *is* $(\log(1/\mu), \mu^{1-c})$*-global with* $\mu \in (0, 1/16)$, *then* $\mathrm{Stab}_\rho(f) \leqslant \mu^{1+c}$.

A key ingredient in the proof is the following lemma proved in the next subsection via global hypercontractivity. Firstly we introduce some notation. Given an orthogonal decomposition $f = \sum_{S\subset[n]} f^{=S}$ and $r \geqslant 0$, we write $f^{\leqslant r} = \sum_{|S|\leqslant r} f^{=S}$ and $f^{>r} = f - f^{\leqslant r}$. We say that f has degree (at most) r if $f = f^{\leqslant r}$.

Lemma 5.15. *For any* $\rho \leqslant 1/80$, *if* $f : \Omega \to \mathbb{R}$ *is* (r, β)*-global of degree* r, *then*

$$\left\|\mathrm{T}_\rho f\right\|_4 \leqslant \beta^{1/4} \|f\|_2^{1/2}.$$

Proof of Theorem 5.14. We start by showing that there exist $\rho', c' > 0$ such that the statement of the theorem holds with (ρ', c') in place of (ρ, c). We take $\rho' = 2^{-200}$ and $c' = 1/100$. Firstly we note that by globalness (applied with no restriction), we have $\mu(f) = \mathbb{E}[f^2] \leqslant \mu^{0.99}$. Let $d =$

$\lfloor c' \log(1/\mu) \rfloor$. We have

$$\mathsf{Stab}_\rho(f) = \sum_{S \subset [n]} \rho^{|S|} \left\| f^{=S} \right\|_2^2 \leqslant \sum_{|S| \leqslant d} \rho^{|S|} \left\| f^{=S} \right\|_2^2 + \rho^{d+1} \left\| f^{>d} \right\|_2^2 = \langle f, T_\rho f^{\leqslant d} \rangle + \rho^{d+1} \left\| f^{>d} \right\|_2^2.$$

Clearly, $\rho^{d+1} \| f^{>d} \|_2^2 \leqslant 2^{-2\log(1/\mu)} = \mu^2$. By Holder's inequality,

$$\langle f, T_\rho f^{\leqslant d} \rangle \leqslant \| f \|_{4/3} \left\| T_\rho f^{\leqslant d} \right\|_4 \leqslant \mu^{0.99} \| f \|_2^{1/2} \leqslant (\mu^{0.99})^{5/4},$$

using Lemma 5.15 and $\| f \|_{4/3} = \mu(f)^{3/4}$ (as f is Boolean), so $\mathsf{Stab}_\rho(f) \leqslant (\mu^{0.99})^{5/4} + \mu^2 \leqslant \mu^{1.01}$.

Now we will deduce the full version of Theorem 5.14, that is, for any $\rho < 1$, there is $c > 0$ such that the statement holds. We let $d = \lceil \log(\rho/\rho') \rceil$, $t = 2^d$ and $c = c'/4t$. By Lemma 5.11, we have $\mathsf{Stab}_\rho(f) \leqslant \| f \|_2^{2(1-1/t)} \mathsf{Stab}_{\rho^t}(f)^{1/t}$. We have $\mathsf{Stab}_{\rho^t}(f) \leqslant \mathsf{Stab}_{\rho'}(f)$ by monotonicity of $\rho \mapsto \mathsf{Stab}_\rho(f)$ and $\rho^t \leqslant \rho'$. By globalness $\mu(f) = \mathbb{E}[f^2] \leqslant \mu^{1-c}$, so $\mathsf{Stab}_\rho(f) \leqslant \mu^{(1-c)(1-1/t)+(1+c')/t} = \mu^{1-c+(c'+c)/t} \leqslant \mu^{1+c}$. □

5.5 | Noise stability via global hypercontractivity

As mentioned in the previous subsection, in this subsection, we will prove the noise stability estimate Lemma 5.15. We start with some definitions required to state our global hypercontractivity inequality. As before, we consider a product measure $\nu = \prod_{i=1}^n \nu_i$ on $\Omega = \prod_{i=1}^n \Omega_i$. For $S \subset [n]$, we let ν_S denote the product measure $\prod_{i \in S} \nu_i$ on $\Omega_S = \prod_{i \in S} \Omega_i$.

Given $f \in L^2(\Omega, \nu)$ with orthogonal decomposition $f = \sum_{S \subset [n]} f^{=S}$ and $T \subset [n]$, the *Laplacian* of f according to T is the function $\mathrm{L}_T f : [m]^n \to \mathbb{R}$ defined by

$$(\mathrm{L}_T f)(x) = \sum_{S \supseteq T} f^{=S}(x).$$

If T is a singleton $\{i\}$, we denote the Laplacian by L_i. We also require the following alternative, more combinatorial, definition of the Laplacian. We let $L_\emptyset$ be the identity operator. For $i \in [n]$, it is easily noted that

$$(\mathrm{L}_i f)(x) = f(x) - \mathop{\mathbb{E}}_{\mathbf{a}_i \sim \nu_i} \left[f(x_1, \dots, x_{i-1}, \mathbf{a}_i, x_{i+1}, \dots, x_n) \right].$$

Then, for $T = \{i_1, \dots, i_d\}$ with $d \geqslant 2$, one can show that L_T may be defined alternatively by composition, that is, $\mathrm{L}_T f = \mathrm{L}_{i_d}(\mathrm{L}_{i_{d-1}}(\dots(\mathrm{L}_{i_1} f)\dots))$. It is not hard to check that this definition does not depend on the order in which the Laplacians are taken and is equivalent to the definition via orthogonal decompositions.

In the next subsection, we will prove the following refined version of the global hypercontractive inequality on product spaces from [16]. For simplicity, we only consider the version required for our purposes, where we bound the 4-norm after applying noise by a function of the 2-norms of the Laplacians.

Theorem 5.16. *Let (Ω,ν) be a finite product space. Then for every $f:\Omega\to\mathbb{R}$ and $\rho\leqslant 1/160$, we have*

$$\left\|\mathrm{T}_\rho f\right\|_4^4 \leqslant \sum_{S\subset[n]} \mathop{\mathbb{E}}_{\mathbf{y}\sim\nu_S}\left[\left\|(\mathrm{L}_S f)_{S\to\mathbf{y}}\right\|_2^4\right].$$

Along with Theorem 5.16, the proof of Lemma 5.15 also requires the following consequence of globalness for norms of Laplacians.

Claim 5.17. Let $f:\Omega\to\mathbb{R}$ be (r,ε)-global, $T\subset[n]$ with $|T|\leqslant r$ and $y\in[m]^T$. Then $\|(\mathrm{L}_T f)_{T\to y}\|_2\leqslant 2^{|T|}\sqrt{\varepsilon}$.

The proof requires the following alternative formula for Laplacians.

Claim 5.18. For any $f:\Omega\to\mathbb{R}$, $T\subset[n]$, we have $(\mathrm{L}_T f)(z)=\sum_{S\subset T}(-1)^{|S|}\mathbb{E}_{\mathbf{a}\sim\nu_S}[f(x_S=\mathbf{a},x_{\overline{S}}=z_{\overline{S}})]$.

Proof. We argue by induction on $|T|$. The claim is immediate from the definition for $|T|=0,1$. Let $|T|=d+1\geqslant 2$, and write $T=T'\cup\{i\}$ with $|T'|=d$. Then by definition and the induction hypothesis,

$$(\mathrm{L}_T f)(z)=\mathrm{L}_i(\mathrm{L}_{T'}f)(z)=\mathrm{L}_i\sum_{S\subset T'}(-1)^{|S|}\mathop{\mathbb{E}}_{\mathbf{a}\sim\nu_S}\left[f(x_S=\mathbf{a},x_{\overline{S}}=z_{\overline{S}})\right].$$

By linearity and the definition of L_i, we deduce

$$(\mathrm{L}_T f)(z)=\sum_{S\subset T'}(-1)^{|S|}\mathop{\mathbb{E}}_{\mathbf{a}\sim\nu_S}\left[f(x_S=\mathbf{a},x_{\overline{S}}=z_{\overline{S}})-\mathop{\mathbb{E}}_{\mathbf{b}\sim\nu_i}\left[f(x_S=\mathbf{a},x_i=\mathbf{b},x_{\overline{S\cup\{i\}}}=z_{\overline{S\cup\{i\}}})\right]\right].$$

The claim follows as $(\mathbf{a},\mathbf{b})$ is distributed according to $\nu_{S\cup\{i\}}$. □

Proof of Claim 5.17. By Claim 5.18 and globalness, we have

$$\begin{aligned}(\mathrm{L}_T f)_{T\to y}(z)&=\sum_{S\subset T}(-1)^{|S|}\mathop{\mathbb{E}}_{\mathbf{a}\sim\nu_S}\left[f(x_S=\mathbf{a},x_{T\setminus S}=y_{T\setminus S},x_{\overline{T}}=z_{\overline{T}})\right]\\&=\sum_{S\subset T}(-1)^{|S|}\mathop{\mathbb{E}}_{\mathbf{a}\sim\nu_S}\left[f_{T\to(\mathbf{a},y_{T\setminus S})}(z_{\overline{T}})\right],\end{aligned}$$

and taking norm over z and using the triangle inequality yields

$$\left\|(\mathrm{L}_T f)_{T\to y}\right\|_2\leqslant\sum_{S\subset T}\mathop{\mathbb{E}}_{\mathbf{a}\sim\nu_S}\left[\left\|f_{T\to(\mathbf{a},y_{T\setminus S})}\right\|_2\right]\leqslant 2^{|T|}\sqrt{\varepsilon}.$$ □

We conclude this subsection with our estimate for noise stability of global functions.

Proof of Lemma 5.15. Suppose $f : \Omega \to \mathbb{R}$ is (r, β)-global of degree r. Let $\rho = 1/80$. By Theorem 5.16,

$$\left\|\mathrm{T}_{\rho/2} f\right\|_4^4 \leqslant \sum_{S \subset [n]} \mathbb{E}_{\mathbf{y} \sim \nu_S} \left[\left\|(\mathrm{L}_S \mathrm{T}_{1/2} f)_{S \to \mathbf{y}}\right\|_2^4\right].$$

By assumption on f, we only need to consider $|S| \leqslant r$, and for such S by Claim 5.17, we have $\|(\mathrm{L}_S f)_{S \to y}\|_2 \leqslant 2^{|S|}\sqrt{\beta}$ for all $y \in \Omega_S$. As $\|(\mathrm{L}_S \mathrm{T}_{1/2} f)_{S \to y}\|_2^2 \leqslant 4^{-|S|}\|(\mathrm{L}_S f)_{S \to y}\|_2^2$, we deduce

$$\left\|\mathrm{T}_{\rho/2} f\right\|_4^4 \leqslant \beta \sum_{S \subset [n]} \mathbb{E}_{\mathbf{y} \sim \nu_S} \left[\left\|(\mathrm{L}_S \mathrm{T}_{1/2} f)_{S \to \mathbf{y}}\right\|_2^2\right].$$

We estimate each summand using Parseval as

$$\mathbb{E}_{\mathbf{y} \sim \nu_S} \left[\left\|(\mathrm{L}_S \mathrm{T}_{1/2} f)_{S \to \mathbf{y}}\right\|_2^2\right] = \sum_{T \supseteq S} 4^{-|T|} \left\|f^{=T}\right\|_2^2 \leqslant \sum_{T \supseteq S} 2^{-|T|} \left\|f^{=T}\right\|_2^2,$$

$$\text{so } \beta^{-1}\left\|\mathrm{T}_{\rho/2} f\right\|_4^4 \leqslant \sum_{S \subset [n]} \sum_{T \supseteq S} 2^{-|T|} \left\|f^{=T}\right\|_2^2 = \sum_{T \subset [n]} \left\|f^{=T}\right\|_2^2 = \|f\|_2^2.$$

□

5.6 | Global hypercontractivity

We conclude this section by proving Theorem 5.16, via our global hypercontractivity inequality from [16]. We start by stating this inequality, for which we require some notation. Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be independent random variables, each with mean 0, variance 1 and $\mathbb{E}[|\mathbf{Z}_i|^4] \leqslant \sigma_i^{-2}$. For $S \subset [n]$, we let $\mathbf{Z}_S = \prod_{i \in S} \mathbf{Z}_i$ and $\sigma_S = \prod_{i \in S} \sigma_i$. We consider multi-linear functions $g(\mathbf{Z}_1, \ldots, \mathbf{Z}_n) = \sum_{S \subset [n]} a_S \mathbf{Z}_S$ with all $a_S \in \mathbb{R}$. For $S \subset [n]$, the discrete derivative of g at S is $\partial_S g(\mathbf{Z}) = \frac{1}{\sigma_S} \sum_{T \supseteq S} a_T \mathbf{Z}_{T \setminus S}$.

Theorem 5.19 Theorem 7.1, [16]. *In the above set up, for $\rho \in [0, 1/16]$, we have $\|\mathrm{T}_\rho g\|_4^4 \leqslant \sum_{S \subset [n]} \sigma_S^2 \|\partial_S g\|_2^4$.*

We will reduce Theorem 5.16 to Theorem 5.19 as follows. Suppose that (Ω, ν) is a product space with $\Omega = [m]^n$ and $f \in L^2(\Omega, \nu)$. We will simulate f via a function $g : \{0,1\}^{nm} \to \mathbb{R}$ which takes nm biased random bits $\{\mathbf{z}_{i,j}\}_{i \in [n], j \in [m]}$, where the bias of $\mathbf{z}_{i,j}$ is $p_{i,j} = \nu_i(j)/4$. Let $\sigma_{i,j} = \sqrt{p_{i,j}(1 - p_{i,j})}$ and $\chi_{i,j}(z_{i,j}) = (z_{i,j} - p)/\sigma_{i,j}$. We note that $\chi_{i,j}$ satisfy the conditions in the above setup, that is, $\mathbb{E}\chi_{i,j} = 0$, $\mathbb{E}\chi_{i,j}^2 = 1$, $\mathbb{E}\chi_{i,j}^4 \leqslant \sigma_{i,j}^{-2}$. For any $S \subset [n]$ and $x \in \Omega_S$, we define the corresponding character $\chi_{S,x} : \{0,1\}^{nm} \to \mathbb{R}$ for $z = (z_{i,j} : i \in [n], j \in [m])$ by setting $\chi_{S,x}(z) = \prod_{i \in S} \chi_{i,x_i}(z_{i,x_i})$; we also write $\sigma_{S,x} = \prod_{i \in S} \sigma_{i,x_i}$. We then define $g : \{0,1\}^{nm} \to \mathbb{R}$ by setting

$$g(z) = \sum_{S \subset [n]} \sum_{x \in \Omega_S} \sigma_{S,x} \left|f^{=S}(x)\right| \chi_{S,x}(z).$$

Claim 5.20. $\|\mathrm{T}_\rho f\|_4^4 \leqslant \|\mathrm{T}_{4\rho} g\|_4^4$.

Proof. Let $\mathcal{S}$ be the set of (S_1, S_2, S_3, S_4) where each $S_\alpha \subset [n]$ and $|\{\alpha : i \in S_\alpha\}| \neq 1$ for all $i \in [n]$. Expanding the definition of the left-hand side, we can write

$$\left\|\mathrm{T}_\rho f\right\|_4^4 = \mathop{\mathbb{E}}_{\mathbf{x}\sim\nu}\left[\sum_{(S_1,S_2,S_3,S_4)\in\mathcal{S}} \rho^{|S_1|+\cdots+|S_4|} f^{=S_1}(\mathbf{x})\cdots f^{=S_4}(\mathbf{x})\right].$$

Also, if $S = (S_1, \ldots, S_4) \in \mathcal{S}$ and $x \in \Omega_{\bigcup S}$, then $\mathbb{E}[\prod_{\alpha=1}^4 \sigma_{S_\alpha,x_{S_\alpha}} \chi_{S_\alpha,x_{S_\alpha}}] \geqslant \prod_{i\in\bigcup S}(p_{i,x_i}/4) = 16^{-|\bigcup S|}\nu_{\bigcup S}(x)$, using $\mathbb{E}[(\sigma_{i,j}\chi_{i,j})^q] \geqslant p_{i,j}/4$ when $q \in \{2, 3, 4\}$, so expanding the right-hand side

$$\left\|\mathrm{T}_{4\rho} g\right\|_4^4 \geqslant \sum_{S=(S_1,S_2,S_3,S_4)\in\mathcal{S}} \sum_{x\in\Omega_S} (4\rho)^{|S_1|+\cdots+|S_4|}\left|f^{=S_1}(x)\right|\cdots\left|f^{=S_4}(x)\right|16^{-|\bigcup S|}\nu_{\bigcup S}(x).$$

As $|\bigcup S| \leqslant (|S_1| + \cdots + |S_4|)/2$, the claim follows. □

To bound $\|\mathrm{T}_{4\rho} g\|_4^4$, we apply (4ρ)-biased hypercontractivity (Theorem 5.19), which is valid if $4\rho \leqslant 1/16$. As $\sigma_{S,x}^2 \leqslant \nu_S(x)$, we get $\|\mathrm{T}_{4\rho} g\|_4^4 \leqslant \sum_{S\subset[n],x\in\Omega_S} \nu_S(x)\|\partial_{(S,x)} g\|_2^4$. For any $S \subset [n]$ and $x \in \Omega_S$, we have

$$\left\|\partial_{(S,x)} g\right\|_2^2 = \frac{1}{\sigma_{S,x}^2}\sum_{T\supseteq S}\sum_{y\in\Omega_{T\setminus S}} \sigma_{(T,x\circ y)}^2 f^{=T}(x,y)^2 \leqslant \sum_{T\supseteq S} \mathop{\mathbb{E}}_{\mathbf{y}\sim\nu_{T\setminus S}}\left[f^{=T}(x,\mathbf{y})^2\right],$$

as $\sigma_{S,x}^{-2}\sigma_{(T,x\circ y)}^2 = \sigma_{T\setminus S,y}^2 \leqslant \nu_{T\setminus S}(y)$. By Fact 5.8 and Parseval ,we get $\|\partial_{(S,x)} g\|_2^2 \leqslant \|(\mathrm{L}_S f)_{S\to x}\|_2^2$, so

$$\left\|\mathrm{T}_\rho f\right\|_4^4 \leqslant \left\|\mathrm{T}_{4\rho} g\right\|_4^4 \leqslant \sum_{S\subset[n],x\in\Omega_S} \nu_S(x)\left\|(\mathrm{L}_S f)_{S\to x}\right\|_2^4 = \sum_{S\subset[n]} \mathop{\mathbb{E}}_{\mathbf{x}\sim\nu_S}\left[\left\|(\mathrm{L}_S f)_{S\to\mathbf{x}}\right\|_2^4\right].$$

This proves Theorem 5.16.

6 | MODERATE ALPHABETS

This section contains the proof of our main result Theorem 1.1 in the case of moderate alphabets, that is, $m > m_0(t)$ is large, but not huge (exponential in n). As discussed previously, the strategy is inspired by that for small m, but we must settle for a regularity lemma (Lemma 5.6) that only provides parts which are uncapturable, so the proof of the junta approximation theorem becomes considerably harder.

As in the case of small m, we want to show that the restrictions defining the regularity decomposition form a t-intersecting family, so we need to find cross-agreements of any fixed size between two pieces of the decomposition. Again we can reduce to finding cross disagreements by taking restrictions, but this reduction is not immediate as with the stronger pseudorandomness condition in the first part, as uncapturability is not preserved by arbitrary restrictions. We therefore start in the first subsection by proving a 'fairness proposition', showing that random restrictions are unlikely to significantly reduce the measure of a code if it is non-negligible (for which the threshold is such that this is only useful when m is not huge). In the second subsection, we then complete the proof of the main theorem for moderate m assuming the junta approximation theorem, and of the junta approximation theorem assuming the existence of fixed cross-agreements between non-negligible uncapturable codes.

The idea for finding cross disagreements is to apply the global small set expansion theorem from the previous section to show that for any code of small measure, we can substantially increase its measure by a combination of taking restrictions and applying a gluing operation, in which we pass to a smaller alphabet by randomly identifying symbols in each coordinate. Here we note that any cross disagreement after gluing must come from a cross disagreement before gluing (this is why we will reduce to disagreements, as we do not have any corresponding statement for finding cross-agreements of some fixed non-zero size). By applying Hoffman's bound to the glued codes rather than the original codes, we thus obtain a much stronger bound on the original measures.

We develop the theory of gluings in the third subsection, which we use for measure boosting in the fourth subsection. We then prove the existence of fixed cross-agreements in the final subsection. The concept of globalness is fundamental throughout, as it is needed for measure boosting, and also to maintain some pseudorandomness condition throughout the repeated restrictions needed for measure boosting. Indeed, as uncapturability is not preserved by arbitrary restrictions, we need a careful combination of taking restrictions and upgrading uncapturability to globalness. We must also take care to remove extraneous agreements that may be introduced by these restrictions, which is possible as globalness implies uncapturability, and the definition of uncapturability is designed for this argument.

6.1 | The fairness proposition

Here we prove the following 'fairness proposition', analogous to that proved for hypergraphs by Keller and Lifshitz [19]. The proofs are quite similar, but we include the details for the convenience of the reader.

Proposition 6.1. *For any $\delta > 0$ and $s \in \mathbb{N}$, there is $C > 0$ such that for any $\mathcal{F} \subset [m]^n$ with $\mu(\mathcal{F}) \geqslant e^{-n/C}$, for uniformly random $\mathbf{S} \in \binom{[n]}{s}$ and $\mathbf{x} \in [m]^{\mathbf{S}}$, we have $\mathbb{P}[\mu(\mathcal{F}_{\mathbf{S}\to\mathbf{x}}) \geqslant (1-\delta)\mu(\mathcal{F})] \geqslant 1-\delta$.*

Proof. Firstly we consider $s = 1$. For each $i \in [n]$, let $V_i = \{a \in [m] \mid \mu(\mathcal{F}_{x_i\to a}) < (1-\delta)\mu(\mathcal{F})\}$.

We suppose for contradiction that the probability of the complementary event is too large, that is, that

$$\Pr_{\mathbf{i}\in[n],\mathbf{a}\in[m]}[\mathbf{a} \in V_{\mathbf{i}}] = \frac{1}{nm}\sum_{i=1}^{n}|V_i| > \delta.$$

Let $I = \{i \in [n] \mid |V_i| \geqslant \frac{\delta}{2}m\}$. We note that $\frac{1}{nm}\sum_{i\in I}|V_i| \geqslant \delta/2$. We consider uniformly random $\mathbf{x} \in [m]^n$ and let $Z = Z(\mathbf{x}) = |\{i \in I : \mathbf{x}_i \in V_i\}|$. Then $Z(\mathbf{x}) = \sum_{i\in I} 1_{\mathbf{x}_i\in V_i}$ is a sum of independent indicator variables with mean

$$\mathbb{E}Z = \sum_{i\in I}|V_i|/m \geqslant \delta n/2.$$

Let $\mathcal{F}'$ be the set of $x \in \mathcal{F}$ such that $|\{i \in I : x_i \in V_i\}| \geqslant (1-\delta/2)\mathbb{E}Z$. By the Chernoff bound, $\mu(\mathcal{F}') \geqslant \mu(\mathcal{F}) - e^{\Omega_\delta(n)} \geqslant (1-\delta/2)\mu(\mathcal{F})$, provided $C = C(\delta, s)$ is sufficiently large.

Now we estimate $E := \mathbb{E}[Z(\mathbf{x})1_{\mathbf{x}\in\mathcal{F}}]$ in two ways. By definition of V_i, we have

$$E = m^{-n}\sum_{x\in\mathcal{F}}\sum_{i\in I}1_{x_i\in V_i} = m^{-n}\sum_{i\in I}\sum_{a\in V_i}|\mathcal{F}_{x_i\to a}| \leqslant m^{-1}\sum_{i\in I}|V_i|(1-\delta)\mu(\mathcal{F}) = (1-\delta)\mu(\mathcal{F})\mathbb{E}Z.$$

On the other hand, by definition of $\mathcal{F}'$, we have

$$E \geqslant m^{-n} \sum_{x \in \mathcal{F}'} (1-\delta/2)\mathbb{E}Z = (1-\delta/2)\mu(\mathcal{F}')\mathbb{E}Z \geqslant (1-\delta/2)^2\mu(\mathcal{F})\mathbb{E}Z.$$

These bounds are contradictory, so the proof for $s = 1$ is complete.

For $s \geqslant 2$, we proceed by induction. We suppose that the statement holds for any $\delta' > 0$ and $s' < s$ with $C = C(\delta', s')$. We let $\delta' = \delta/2$ and $s' = s-1$ and consider uniformly random $\mathbf{S}' \in \binom{[n]}{s'}$ and $\mathbf{x}' \in [m]^{\mathbf{S}'}$. By the induction hypothesis, which can be applied if we choose $C(\delta, s) > C(\delta', s')$, we have $\mathbb{P}[E_1(\mathbf{S}', \mathbf{x}')] \geqslant 1-\delta'$, where $E_1(\mathbf{S}', \mathbf{x}')$ is the event that $\mu(\mathcal{F}_{\mathbf{S}' \to \mathbf{x}'}) \geqslant (1-\delta')\mu(\mathcal{F})$.

For each S', x' such that $E_1(S', x')$ holds, we consider $\boldsymbol{S} = S' \cup \{\boldsymbol{i}\}$ and $\mathbf{x} = (x', \mathbf{a}) \in [m]^{\mathbf{S}}$ for uniformly random $\mathbf{i} \in [n] \setminus S'$ and $\mathbf{a} \in [m]$. We have $\mu(\mathcal{F}_{S' \to x'}) \geqslant (1-\delta')\mu(\mathcal{F}) > e^{-(n-s+1)/C(\delta', 1)}$ for large $C(\delta, s)$. Applying the base case to $\mathcal{F}_{S' \to x'}$, we have $\mathbb{P}[E_2(\mathbf{S}, \mathbf{x})] \geqslant 1-\delta'$, where $E_2(\mathbf{S}, \mathbf{x})$ is the event that $\mu(\mathcal{F}_{\mathbf{S} \to \mathbf{x}}) \geqslant (1-\delta')\mu(\mathcal{F}_{S' \to x'})$. With probability at least $(1-\delta')^2 \geqslant 1-\delta$ both E_1 and E_2 hold, and we then have $\mu(\mathcal{F}_{S \to x}) \geqslant (1-\delta')^2\mu(\mathcal{F}) \geqslant (1-\delta)\mu(\mathcal{F})$, as required. □

6.2 | Proof summary

In this subsection, we complete the proof of the main theorem for moderate m assuming the junta approximation theorem, and of the junta approximation theorem assuming the existence of fixed cross-agreements between non-negligible uncapturable codes. As m is large, the largest ball is a subcube of co-dimension t, so we can restate our main result for moderate m as follows.

Theorem 6.2. *For any $t \in \mathbb{N}$, there are $m_0, N \in \mathbb{N}$ such that if $m \geqslant m_0$, $n \geqslant N \log m$ and $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding, then $|\mathcal{F}| \leqslant m^{n-t}$, with equality only when $\mathcal{F}$ is a subcube of co-dimension t.*

We will prove Theorem 6.2 assuming the following junta approximation theorem.

Theorem 6.3. *For every $t, k \in \mathbb{N}$, there exist $C, m_0, N \in \mathbb{N}$ such that if $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding with $m \geqslant m_0$ and $n \geqslant N \log m$, then there is a t-intersecting collection $\mathcal{D}$ of at most C subcubes of co-dimension at most k such that $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant Cm^{-k}$.*

Proof of Theorem 6.2. Suppose $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding with $\mu(\mathcal{F}) \geqslant m^{-t}$. By Theorem 6.3, there is a t-intersecting collection $\mathcal{D}$ of $O_t(1)$ subcubes of co-dimension at most $t+1$ such that $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant O_t(1)m^{-(t+1)}$. As $\mathcal{D}$ is t-intersecting, its subcubes all have co-dimension at least t. Let $\mathcal{D}'$ consist of the subcubes in $\mathcal{D}$ that have co-dimension t. Then $\mu(\bigcup \mathcal{D} \setminus \bigcup \mathcal{D}') \leqslant O_t(1)m^{-(t+1)}$. As $O_t(1)m^{-(t+1)} < m^{-t} \leqslant \mu(\mathcal{F})$ for large m, we must have $\mathcal{D}' \neq \emptyset$. Thus, $\mathcal{D}'$ consists of exactly one subcube of co-dimension t, say $\mathcal{S} = \{x \in [m]^n \mid x_1 = 1, \ldots, x_t = 1\}$.

Write $\mu(\mathcal{F}_{[t] \to 1}) = 1-\varepsilon$, where $0 \leqslant \varepsilon = m^t\mu(\mathcal{S} \setminus \mathcal{F}) \leqslant m^t\mu(\mathcal{F} \setminus \mathcal{S}) \leqslant O_t(m^{-1})$. Suppose for contradiction $\varepsilon > 0$. We claim that $\varepsilon > e^{-2tn/N}$. To see this, fix any $a \in \mathcal{F} \setminus \mathcal{S}$ (using $\varepsilon > 0$). Write $|\{i \in [t] : a_i = 1\}| = t-1-s$ with $s \geqslant 0$, fix any $S \subset [n] \setminus [t]$ with $|S| = s$, and let $R = [n] \setminus ([t] \cup S)$. For $b = a_S$ and $c \in [m]^R$ with $\mathrm{agr}(c, a_R) = 0$, we have $(1^t, b, c) \notin \mathcal{F}$ (since $(1^t, b, c)$ and a agree on $t-1$ coordinates), giving $|\mathcal{S} \setminus \mathcal{F}| \geqslant (m-1)^{n-t-s}$, so $\varepsilon \geqslant m^{-s}(1-1/m)^{n-t-s} > e^{-tn/N-2n/m} \geqslant e^{-2tn/N}$, as claimed.

As $\mu(\mathcal{F} \setminus \mathcal{S}) \geqslant m^{-t}\varepsilon$, by averaging, we can fix $1^t \neq x \in [m]^t$ with $\mu(\mathcal{F}_{[t] \to x}) \geqslant m^{-t}\varepsilon$. Write $|\{i \in [t] : a_i = 1\}| = t-1-s$ with $s \geqslant 0$. Consider uniformly random $\mathbf{S} \in \binom{[n] \setminus [t]}{s}$ and $\mathbf{y} \in [m]^{\mathbf{S}}$. Let

$\mathcal{G} = \mathcal{F}_{[t]\to 1, \mathbf{S}\to \mathbf{y}}$ and $\mathcal{H} = \mathcal{F}_{[t]\to x, \mathbf{S}\to \mathbf{y}}$. By Markov's inequality, $\mathbb{P}[\mu(\mathcal{G}) \geqslant 1 - 2\varepsilon] \geqslant 1/2$. By Proposition 6.1, $\mathbb{P}[\mu(\mathcal{H}) \geqslant 0.9m^{-t}\varepsilon] \geqslant 0.9$. Thus, we can fix (S, x) so that $\mu(\mathcal{G}) \geqslant 1 - 2\varepsilon$ and $\mu(\mathcal{H}) \geqslant 0.9m^{-t}\varepsilon$. However, $\mathcal{G}$ and $\mathcal{H}$ are cross intersecting, so this contradicts Theorem 2.3. Thus, $\varepsilon = 0$, as required. □

We conclude this subsection by proving Theorem 6.3 assuming the following result on cross-agreements between uncapturable codes, the proof of which will be the goal of the remainder of this section.

Theorem 6.4. *For any $s, k \in \mathbb{N}$, there are $r, m_0, N \in \mathbb{N}$ such that if $m \geqslant m_0$, $n \geqslant N \log m$ and $\mathcal{A}_j \subset [m]^{[n]\setminus R_j}$ are (r, m^{-k})-uncapturable with $|R_j| \leqslant k$ for $j = 1, 2$ then there are $x^j \in \mathcal{A}_j$ for $j = 1, 2$ with $|\{i \in [n] \setminus (R_1 \cup R_2) : x^1_i = x^2_i\}| = s$.*

Proof of Theorem 6.3. Suppose $r, m, N \gg t, k$ and let $\mathcal{F} \subset [m]^n$ with $n \geqslant N \log m$ be $(t-1)$-avoiding. By Lemma 5.6 with $\varepsilon = 1$, there is a collection $\mathcal{D}$ of at most r^k subcubes of co-dimension at most k such that $\mathcal{F}_{R\to\alpha}$ is (r, m^{-k})-uncapturable for each $D = D_{R\to\alpha} \in \mathcal{D}$ and $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant 3r^{k+1}m^{-k}$. Suppose for a contradiction that $\mathcal{D}$ is not t-intersecting. Then there are $D_{R^j\to\alpha^j} \in \mathcal{D}$ for $j = 1, 2$ (not necessarily different) that agree on $t - 1 - s$ coordinates for some $s \geqslant 0$. Let $\mathcal{A}_1 = \mathcal{F}_{R^1\to\alpha^1} \setminus \bigcup_{i\in R^2\setminus R^1} D_{i\to\alpha^2_i} \subset [m]^{[n]\setminus R^1}$ and define $\mathcal{A}_2$ similarly. Then $\mathcal{A}_1$, $\mathcal{A}_2$ are $(r-k, m^{-k})$-uncapturable, so by Theorem 6.4, there are $x^j \in \mathcal{A}_j$ with $|\{i \in [n] \setminus (R^1 \cup R^2) : x^1_i = x^2_i\}| = s$. But then $\mathsf{agr}((\alpha^1, x^1), (\alpha^2, x^2)) = t - 1$, which is a contradiction. □

6.3 | Gluings and expansion

In this subsection, we introduce our gluing operation and establish a small set expansion property for global codes under random gluings.

Definition 6.5. Let $k < m \in \mathbb{N}$ and $b \geqslant 1$. A b-balanced gluing from $[m]$ to $[k]$ is a function $\pi : [m] \to [k]$ such that $|\pi^{-1}(i)| \leqslant bm/k$ for all $i \in [k]$. We let $\Pi_{m,k,b}$ denote the set of all such gluings. If $b = 1$ (which is only possible when $k \mid m$), we may omit it from our notation.

A b-balanced gluing of $[m]^n$ to $[k]^n$ is a mapping $\pi : [m]^n \to [k]^n$ of the form $\pi(x_1, \dots, x_n) = (\pi_1(x_1), \dots, \pi_n(x_n))$ with $\pi_1, \dots, \pi_n \in \Pi_{m,k,b}$. We let $\Pi^{\otimes n}_{m,k,b}$ denote the set of all such gluings; we may omit the superscript if n is clear from context. For $\mathcal{F} \subset [m]^n$ and $\pi \in \Pi^{\otimes n}_{m,k,b}$, we write $\mathcal{F}^\pi = \pi(\mathcal{F}) \subset [k]^n$.

Example 6.6. Consider the gluing $\pi : [3]^n \to [2]^n$ where for each $i \in [n]$, we have $\pi_i(1) = \pi_i(2) = 1$ and $\pi_i(3) = 2$. Let $\mathcal{F} = \{x \in [3]^n \mid |\{i \mid x_i = 1 \vee x_i = 2\}| \geqslant \frac{2}{3}n\}$. Then $\mathcal{F}$ has constant measure in $[3]^n$, but $\mathcal{F}^\pi$ has exponentially small measure in $[2]^n$.

This example indicates that we should make a careful choice of measure in $[k]^n$ for gluing to be useful.

Definition 6.7. Given a measure ν on $[m]$ and $\pi : [m] \to [k]$, we define a measure ν^π on $[k]$ by $\nu^\pi(x) = \sum_{y\in\pi^{-1}(x)} \nu(y)$. Given a product measure $\nu = \prod_{i=1}^n \nu_i$ on $[m]^n$ and $\pi = (\pi_1, \dots, \pi_n)$

with each $\pi_i : [m] \to [k]$, we define a product measure $\nu^\pi = \prod_{i=1}^n \nu_i^\pi$ on $[k]^n$ by $(\nu^\pi)_i = (\nu_i)^{\pi_i} = \sum_{y \in \pi_i^{-1}(x)} \nu(y)$ for each i. We say that ν is b-balanced if $\nu_i(x) \leqslant b/m$ for all $i \in [n]$ and $x \in [m]$.

Claim 6.8. With notation as in Definition 6.7, for any $\mathcal{F} \subset [m]^n$, we have $\nu^\pi(\mathcal{F}^\pi) \geqslant \nu(\mathcal{F})$.

Proof. For any $y \in [k]^n$, we have

$$\nu^\pi(y) = \prod_{i=1}^n \nu_i^\pi(y_i) = \prod_{i=1}^n \sum_{x_i \in \pi_i^{-1}(y_i)} \nu_i(x_i) = \sum_{x \in \pi^{-1}(y)} \prod_{i=1}^n \nu_i(x_i) = \sum_{x \in \pi^{-1}(y)} \nu(x), \quad \text{so}$$

$$\nu^\pi(\mathcal{F}^\pi) = \sum_{y \in [k]^n} \nu_\pi(y) 1_{y \in \mathcal{F}^\pi} = \sum_{y \in [k]^n} \sum_{x \in \pi^{-1}(y)} \nu(x) 1_{y \in \mathcal{F}^\pi} \geqslant \sum_{y \in [k]^n} \sum_{x \in \pi^{-1}(y)} \nu(x) 1_{x \in \mathcal{F}} = \nu(\mathcal{F}).$$

□

Now we establish global small set expansion for random balanced gluings.

Lemma 6.9. *With notation as in Definitions 6.5 and 6.7, there is $c > 0$ such that the following holds. Let $s, k, m \in \mathbb{N}$ be such that $k = m/s$ and $s \geqslant 4$, let ν be an s-balanced product measure on $[m]^n$, and suppose $\mathcal{F} \subset [m]^n$ is $(\log(1/\mu), \mu^{1-c})$-global with $\mu \in (0, 1/16)$. Then $\mathbb{E}_{\pi \in \Pi_{m,k}^{\otimes n}}[\nu^\pi(\mathcal{F}^\pi)] \geqslant \nu(\mathcal{F})^{1-c}$.*

Proof. The plan for the proof is to show $\mathbb{E}_\pi[\nu^\pi(\mathcal{F}^\pi)] \geqslant \nu(\mathcal{F})^2/\langle f, Tf\rangle$, where f is the characteristic function of $\mathcal{F}$ and $T = \prod_{i=1}^n T_i$ is some product Markov chain on $[m]^n$ with each $\lambda_*(T_i) \geqslant 1/6$. By Theorem 5.3, this will suffice to establish the lemma.

To construct T, we first consider for each π the operator $T_\pi^\uparrow : L^2([m]^n, \nu) \to L^2([k]^n, \nu^\pi)$ defined by $T_\pi^\uparrow f(y) = \mathbb{E}_{\mathbf{x} \sim \nu}[f(\mathbf{x}) \mid \pi(\mathbf{x}) = y]$ for any $y \in [k]^n$. Note that $\nu(\mathcal{F}) = \nu(f) = \nu^\pi(T_\pi^\uparrow f)$, as if $y \sim \nu^\pi$ and $x \sim \nu \mid \pi(x) = y$, then $x \sim \nu$. Writing f^π for the characteristic function of $\mathcal{F}^\pi$, by Cauchy–Schwarz, we can bound $\nu(\mathcal{F})^2 = \mathbb{E}_\pi[\nu^\pi(T_\pi^\uparrow f)]^2$ as

$$\mathop{\mathbb{E}}_\pi\left[\nu^\pi(T_\pi^\uparrow f)\right]^2 = \mathop{\mathbb{E}}_\pi\left[\langle T_\pi^\uparrow f, f^\pi\rangle_{\nu^\pi}\right]^2 \leqslant \mathop{\mathbb{E}}_\pi\left[\left\|T_\pi^\uparrow f\right\|_{2,\nu^\pi} \|f^\pi\|_{2,\nu_\pi}\right]^2 \leqslant \mathop{\mathbb{E}}_\pi\left[\left\|T_\pi^\uparrow f\right\|_{2,\nu^\pi}^2\right] \mathop{\mathbb{E}}_\pi\left[\|f^\pi\|_{2,\nu^\pi}^2\right].$$

We note that $\mathbb{E}_\pi[\|f^\pi\|_{2,\nu^\pi}^2] = \mathbb{E}_\pi[\nu^\pi(\mathcal{F}^\pi)]$ is the expression that we wish to bound. We write

$$\mathop{\mathbb{E}}_\pi\left[\left\|T_\pi^\uparrow f\right\|_{2,\nu_\pi}^2\right] = \mathop{\mathbb{E}}_{\substack{\pi \\ \mathbf{y} \sim \nu_\pi \\ \mathbf{x}, \mathbf{x}' \sim \nu}}\left[f(\mathbf{x}) f(\mathbf{x}') \mid \boldsymbol{\pi}(\mathbf{x}) = \boldsymbol{\pi}(\mathbf{x}') = \mathbf{y}\right] = \langle f, Tf\rangle,$$

where T is the reversible Markov chain on $[m]^n$ characterised by the property that two consecutive states $\mathbf{x}, \mathbf{x}'$ of its stationary chain are distributed as independent samples from ν conditioned on $\boldsymbol{\pi}(\mathbf{x}) = \boldsymbol{\pi}(\mathbf{x}') = \mathbf{y}$, where $\pi \sim \Pi_{m,k}^{\otimes n}$ and $y \sim \nu^\pi$. We note that each of $\mathbf{x}, \mathbf{x}'$ then has marginal distribution ν, which is therefore the stationary distribution. As coordinates are independent, we can write $T = \prod_{i=1}^n T_i$ as a product chain. To complete the proof, it remains to show each $\lambda_*(T_i) \geqslant 1/6$. By Lemma 5.7, it suffices to prove the following claim.

Claim 6.10. For any $i \in [n]$ and $a, b \in [m]$, we have $p_i(a, b) := \mathbb{P}(\mathbf{x}_i = a, \mathbf{x}_i' = b) \geqslant \frac{1}{6}\nu_i(a)\nu_i(b)$.

To see this, we expand out the definition to write

$$p_i(a,b) = \mathop{\mathbb{E}}_{\pi}\left[\sum_{j\in[k]} \nu_i^\pi(j) 1_{\pi(a)=\pi(b)=j} \frac{\nu_i(a)}{\nu_i^\pi(j)} \frac{\nu_i(b)}{\nu_i^\pi(j)}\right] = \nu_i(a)\nu_i(b) \sum_{j\in[k]} \mathop{\mathbb{E}}_{\pi}\left[1_{\pi(a)=\pi(b)=j} \frac{1}{\nu_i^\pi(j)}\right].$$

Each $\mathbb{P}(\pi(a) = \pi(b) = j) = \frac{1}{k}\frac{s-1}{m-1} \geqslant \frac{1}{2k^2}$, so by Jensen's inequality

$$p_i(a,b) \geqslant \frac{\nu_i(a)\nu_i(b)}{2k^2} \sum_{j\in[k]} \mathop{\mathbb{E}}_{\pi}\left[\frac{1}{\nu_i^\pi(j)} \,\middle|\, {\pi(a)=j, \atop \pi(b)=j}\right] \geqslant \frac{\nu_i(a)\nu_i(b)}{2k^2} \sum_{j\in[k]} \frac{1}{\mathbb{E}_\pi\left[\nu_i^\pi(j) \,\middle|\, \pi(a)=\pi(b)=j\right]}.$$

As $\pi^{-1}(j)$ consists of a, b and $s-2$ uniformly random elements from $[m] \setminus \{a,b\}$, we have

$$\mathop{\mathbb{E}}_{\pi}\left[\nu_i^\pi(j) \,\middle|\, \pi(a)=\pi(b)=j\right] = \nu_i(a) + \nu_i(b) + \frac{s-2}{m-2}\sum_{x\neq a,b} \nu_i(x) \leqslant \nu_i(a) + \nu_i(b) + \frac{s}{m} \leqslant \frac{3}{k},$$

as each $\nu_i(y) \leqslant s/m = 1/k$. Thus, $p_i(a,b) \geqslant \frac{1}{2k^2}\nu_i(a)\nu_i(b)\sum_{j\in[k]}\frac{k}{3} = \frac{1}{6}\nu_i(a)\nu_i(b)$. This completes the proof of the claim, and so of the lemma. □

6.4 | Boosting measure

In this subsection, we apply the small set expansion properties of random gluings established in the previous subsection to prove the following result, which shows that the measure of any small code can be substantially increased via restrictions and gluings.

Lemma 6.11. *For every $\varepsilon > 0$, there is $C > 0$ such that for any b-balanced product measure ν on $[m]^n$ with $4 \leqslant b \in \mathbb{N}$ and $m > b^{3C}$, if $\mathcal{F} \subset [m]^n$ with $\nu(\mathcal{F}) = \mu < 16^{-1/\varepsilon}$, then there are $\pi \in \Pi_{m,m',b}$ with $m' > m/b^{2C+1}$ and $\alpha \in [m']^R$, where $R \subset [n]$ with $|R| < C\log(\mu^{-1})$, such that $\nu^\pi((\mathcal{F}^\pi)_{R\to\alpha}) \geqslant \mu^\varepsilon$.*

Proof. We start by applying an arbitrary b-balanced gluing $\pi_0 \in \Pi^{\otimes n}_{m,m_0,b}$, where m_0 is the largest power of b that is at most m. Clearly, $\nu_0 := \nu^{\pi_0}$ is b^2-balanced. We let $\mathcal{F}_0 = \mathcal{F}^{\pi_0} \subset [m_0]^{S_0}$, where $S_0 = [n]$. By Claim 6.8, we have $\mu_0 := \nu_0(\mathcal{F}_0) \geqslant \mu$.

Now we apply the following iterative procedure for $i \geqslant 0$. Given $\mathcal{F}_i \subset [m_i]^{S_i}$, where $S_0 = [n]$, with $\nu_i(\mathcal{F}_i) = \mu_i \geqslant \mu$ and ν_i is a b^2-balanced product measure,

(1) if $\mu_i \geqslant \mu^\varepsilon$ we stop, otherwise,
(2) if $\mathcal{F}_i$ is not $(\log(1/\mu_i), \mu_i^{1-c})$-global according to ν_i, where $c > 0$ is as in Lemma 6.9, then by definition, we can choose $\mathcal{F}_{i+1} = (\mathcal{F}_i)_{R_i\to\alpha_i} \subset [m_{i+1}]^{S_{i+1}}$ with $\mu_{i+1} = \nu_{i+1}(\mathcal{F}_{i+1}) \geqslant \mu_i^{1-c}$, where $m_{i+1} = m_i$, $\nu_{i+1} = \nu_i$ and $S_{i+1} = S_i \setminus R_i$ for some R_i with $|R_i| \leqslant \log(1/\mu_i)$ and $\alpha_i \in [m_i]^{R_i}$,
(3) otherwise, as $\mu_i < \mu^\varepsilon \leqslant 1/16$, by Lemma 6.9, we can choose $\mathcal{F}_{i+1} = (\mathcal{F}_i)^{\pi_i} \subset [m_{i+1}]^{S_{i+1}}$ with $m_{i+1} = m_i/b^2$, $S_{i+1} = S_i$, $\pi \in \Pi_{m_i,m_{i+1}}$ and $\mu_{i+1} = \nu_{i+1}(\mathcal{F}_{i+1}) \geqslant \mu_i^{1-c}$, where $\nu_{i+1} = \nu_i^{\pi_i}$.

If $C > C_0(\varepsilon, c)$ is large, then this process terminates in at most C steps, with some $\mathcal{F}_r \subset [m_r]^{S_r}$, where $m_r \geqslant m/(b^{2C+1})$ and $S_r = [n] \setminus R$, where R is the union of all sets R_i in the process, so $|R| \leqslant C\log(1/\mu)$. For $i \geqslant 0$, we let $\pi_{i\to r} \in \Pi_{m_i,m_r}$ be obtained by composing all π_j with $i < j \leqslant r$. We define $\alpha \in [m_r]^R$ by $\alpha_x = \pi_{i\to r}((\alpha_i)_x)$ for $x \in R_i$. We let $\pi = \pi_{0\to r}$ and note that $\nu^\pi = \nu_r$ and $\mathcal{F}_r \subset (\mathcal{F}^\pi)_{R\to\alpha}$, so $\nu^\pi((\mathcal{F}^\pi)_{R\to\alpha}) \geqslant \nu_r(\mathcal{F}_r) \geqslant \mu^\varepsilon$. □

6.5 | Uncapturable codes agree

In this subsection, we prove our cross-agreement result for uncapturable codes, Theorem 6.4. As demonstrated in Subsection 6.2, this will complete the proof of our main theorem for moderate alphabets. We start with an outline of the proof. We are given two uncapturable codes $\mathcal{A}_1$ and $\mathcal{A}_2$ and need to find a cross-agreement of some fixed size s. Moreover, the coordinate sets may be slightly different: we have $\mathcal{A}_j \subset [m]^{[n]\setminus R_j}$ with $|R_j| \leqslant k$ for $j = 1, 2$.

Step 1: Globalness. We would like to restrict to a common coordinate set, but we cannot do so immediately, as uncapturability is not closed under restrictions. We therefore start by upgrading to globalness, while avoiding unwanted agreements. We find a global code $\mathcal{A}_1'$ obtained from $\mathcal{A}_1$ by a small restriction. We obtain $\mathcal{A}_2'$ from $\mathcal{A}_2$ by removing any agreements with this restriction, using uncapturability to see that $\mathcal{A}_2'$ is not negligible, and find a global code $\mathcal{B}_2$ obtained from $\mathcal{A}_2'$ by a small restriction. Then we obtain a global code $\mathcal{B}_1$ from $\mathcal{A}_1'$ by removing any agreements with this restriction.

Step 2: Fairness. By the fairness proposition, we find a common restriction of size s by which we obtain non-negligible global codes $\mathcal{C}_1, \mathcal{C}_2$ from $\mathcal{B}_1, \mathcal{B}_2$. It remains to show that $\mathcal{C}_1, \mathcal{C}_2$ cannot be cross-agreeing.

Step 3: Expansion. We apply measure boosting to find a gluing and restriction so that $\mathcal{C}_2$ becomes some $\mathcal{C}_2'$ with dramatically larger measure. We obtain $\mathcal{C}_1'$ from $\mathcal{C}_1$ by removing any extra agreements created by the gluing and restriction, and then $\mathcal{C}_1''$ with non-negligible measure by applying the gluing that was found for $\mathcal{C}_2$. We now find a gluing and restrictions for $\mathcal{C}_1''$ to get from it a family $\mathcal{C}_1'''$ with dramatically larger measure than $\mathcal{C}_1''$. We then remove these restrictions as well as apply this gluing on $\mathcal{C}_2'$ to get $\mathcal{C}_2'''$ whose measure not much smaller than that of $\mathcal{C}_2'$. By averaging, we can apply further restrictions without reducing measures to obtain $\mathcal{G}_1, \mathcal{G}_2$ on a common set of coordinates.

Step 4: Hoffman bound. The measures of $\mathcal{G}_1, \mathcal{G}_2$ are so large that they cannot be cross-agreeing, so we find a cross disagreement, which corresponds to an agreement of size s in the original codes.

We proceed to the formal proof of Theorem 6.4.

Proof of Theorem 6.4. We are given (r, m^{-k})-uncapturable $\mathcal{A}_j \subset [m]^{[n]\setminus R_j}$ with $|R_j| \leqslant k$ for $j = 1, 2$, and we need to find $x^j \in \mathcal{A}_j$ with $|\{i \in [n] \setminus (R_1 \cup R_2) : x_i^1 = x_i^2\}| = s$, where $n \geqslant N \log m$ and $r, m, N \gg s, k$.

Step 1: Globalness. By uncapturability $\mu(\mathcal{A}_1) \geqslant m^{-k}$, so by Lemma 5.5 with $\gamma = m^{-1/10}$ and $r/100k$ in place of r, we obtain $\mathcal{A}_1' = (\mathcal{A}_1)_{R_1' \to \alpha_1'}$ that is $(r/100k, \mu(\mathcal{A}_1')/\gamma)$-global with $\mu(\mathcal{A}_1') \geqslant \mu(\mathcal{A}_1)$, where $|R_1'| \leqslant \log_{1/\gamma}(1/\mu(\mathcal{A}_1))r/100k \leqslant r/10$. We note that $\mathcal{A}_2' := \mathcal{A}_2 \setminus \bigcup_{i \in R_1'} D_{i \to \alpha_1'(i)}$ is $(0.9r, m^{-k})$-uncapturable, so $\mu(\mathcal{A}_2') \geqslant m^{-k}$. From Lemma 5.5, we obtain $\mathcal{B}_2 = (\mathcal{A}_2')_{R_2' \to \alpha_2'}$ that is $(r/100k, \mu(\mathcal{B}_2)/\gamma)$-global with $\mu(\mathcal{B}_2) \geqslant \mu(\mathcal{A}_2')$, where $|R_2'| \leqslant r/10$. In particular, $\mathcal{B}_2 \neq \emptyset$, so $R_2' \to \alpha_2'$ has no agreement with $R_1' \to \alpha_1'$. We let $\mathcal{B}_1 = \mathcal{A}_1' \setminus \bigcup_{i \in R_2'} D_{i \to \alpha_2'(i)}$. By Claim 5.4, $\mathcal{A}_1'$ is $(\gamma m/4, \mu(\mathcal{A}_1')/2)$-uncapturable, so $\mu(\mathcal{B}_1) \geqslant \frac{1}{2}\mu(\mathcal{A}_1')$, which implies that $\mathcal{B}_1$ is $(r/100k, 2\mu(\mathcal{B}_1)/\gamma)$-global.

Step 2: Fairness. As $n \geqslant N \log m$ and N is large, we have $\mu(\mathcal{B}_1), \mu(\mathcal{B}_2) \geqslant \frac{1}{2}m^{-k} \geqslant e^{-n/C}$, where $C = C(s, 0.1)$ is as in Proposition 6.1. Consider uniformly random $\mathbf{S} \subset [n] \setminus (R_1 \cup R_1' \cup R_2 \cup R_2')$ of size s and $\mathbf{z} \in [m]^{\mathbf{S}}$. For large n, the distribution of $\mathbf{S}$ has total variation distance $o(1)$ from the uniform distribution on $\binom{[n]\setminus(R_1 \cup R_1')}{s}$. Thus, by Proposition 6.1, we have $\mathbb{P}[\mu((\mathcal{B}_1)_{\mathbf{S}\to\mathbf{z}}) \geqslant 0.9\mu(\mathcal{B}_1)] \geqslant$

$0.9 - o(1)$, and similarly for $\mathcal{B}_2$. Thus, we can fix S and z so that both $\mathcal{C}_j = (\mathcal{B}_j)_{S\to z}$ have $\mu(\mathcal{C}_j) \geqslant \frac{1}{2}\mu(\mathcal{B}_j)$, so are $(r/100k, 4\mu(\mathcal{C}_j)/\gamma)$-global.

Step 3: Expansion. By Lemma 6.11 applied to $\mathcal{C}_2$ with $\varepsilon = 1/3k$ and $b = 4$, there are $\pi_2 \in \Pi_{m,m_2,4}$ with $m_2 = \Omega_k(m)$, $\alpha_2'' \in [m_2]^{R_2''}$, where $R_2'' \subset [n] \setminus (R_2 \cup R_2' \cup S)$ with $|R_2''| < O_k(\log m) \ll n$, such that $\mathcal{C}_2' := (\mathcal{C}_2^{\pi_2})_{R_2''\to\alpha_2''}$ has $\mu^{\pi_2}(\mathcal{C}_2') \geqslant 1/\sqrt{m}$. Let

$$\mathcal{C}_1' = \mathcal{C}_1 \setminus \bigcup \{D_{i\to a} : i \in R_2'', (\pi_2)_i(a) = (\alpha_2'')_i\}.$$

By Claim 5.4, $\mathcal{C}_1$ is $(\gamma m/16, \mu(\mathcal{C}_1)/2)$-uncapturable, so $\mu(\mathcal{C}_1') \geqslant \frac{1}{2}\mu(\mathcal{C}_1)$. Let $\mathcal{C}_1'' = (\mathcal{C}_1')^{\pi_2}$. By Claim 6.8, we have $\mu^{\pi_2}(\mathcal{C}_1'') \geqslant \mu(\mathcal{C}_1') \geqslant \frac{1}{8}m^{-k}$.

By Lemma 6.11 applied to $\mathcal{C}_1''$ under the 4-balanced measure μ^{π_2} with $\varepsilon = 1/3k$ and $b = 4$, there are $\pi_1 \in \Pi_{m_2,m_1,4}$ with $m_1 = \Omega_k(m)$, $\alpha_1'' \in [m_1]^{R_1''}$, where $R_1'' \subset [n] \setminus (R_1 \cup R_1' \cup S)$ with $|R_1''| < O_k(\log m) \ll n$, such that $\mathcal{C}_1''' := ((\mathcal{C}_1'')^{\pi_1})_{R_1''\to\alpha_1''}$ has $\mu^{\pi_1\circ\pi_2}(\mathcal{C}_1''') \geqslant 1/\sqrt{m}$. Let

$$\mathcal{C}_2'' = \mathcal{C}_2' \setminus \bigcup \{D_{i\to a} : i \in R_1'', (\pi_1)_i(a) = (\alpha_1'')_i\}.$$

Then $\mu^{\pi_2}(\mathcal{C}_2'') \geqslant \mu(\mathcal{C}_2') - O_k(m^{-1}\log m) \geqslant 1/2\sqrt{m}$. Let $\mathcal{C}_2''' = (\mathcal{C}_2'')^{\pi_1}$. By Claim 6.8, we have $\mu^{\pi_1\circ\pi_2}(\mathcal{C}_2''') \geqslant \mu(\mathcal{C}_2'') \geqslant 1/2\sqrt{m}$.

Step 4: Hoffman bound. By averaging, we can choose restrictions $\mathcal{G}_j \subset [m_1]^{[n]\setminus R}$ of $\mathcal{C}_j'''$ for $j = 1, 2$ where $R = R_1 \cup R_1' \cup R_1'' \cup R_2 \cup R_2' \cup R_2'' \cup S$ such that both $\nu(\mathcal{G}_j) \geqslant \nu(\mathcal{C}_j''') \geqslant 1/2\sqrt{m}$, where $\nu = \mu^{\pi_1\circ\pi_2}$ is 16-balanced. By construction, the elements of $\mathcal{G}_j$ for $j = 1, 2$ are of the form $\pi_1\pi_2(x^j_{[n]\setminus R})$ where $x^j \in \mathcal{A}_j$ with $|\{i \in R \setminus (R_1 \cup R_2) : x_i^1 = x_i^2\}| = s$. By Lemma 5.9 applied with $\lambda = O_k(1/m)$, we can find a cross disagreement, which corresponds to $x^j \in \mathcal{A}_j$ with $|\{i \in [n] \setminus (R_1 \cup R_2) : x_i^1 = x_i^2\}| = s$. $\square$

Part III: Huge alphabets

This part contains the proof of our main result Theorem 1.1 in the case of huge alphabets, that is, when $n \leqslant N(t)\log m$, with $N(t)$ as in Theorem 6.2. As previously discussed, there are examples showing that a proof strategy based on cross agreements between uncapturable codes cannot work in this setting, so instead we adopt a combinatorial perspective.

7 | FORBIDDEN AGREEMENT CONFIGURATIONS

The proofs in this section exploit combinatorial arguments that obtain expansion in measure from a 'shadow' operation, which is analogous (but quite different in various details) to an argument in the hypergraph setting due to Keller and Lifshitz [19]. This operation requires us to consider more general agreement configurations (which are anyway of interest) even if we only want to find pairwise agreements as in our main result. We introduce these configurations and their interpretation in terms of expanded hypergraphs in the first subsection, and prove an extremal result for configurations. In the second subsection, we define our shadow operation and establish two key properties, namely that a code with a given forbidden configuration (a) has an average shadow with much larger measure and (b) there is some shadow with much stronger uncapturability.

We extend these properties in the third subsection to iterated shadows when there is some forbidden configuration with a 'kernel', that is, some common intersection of all restrictions in the configuration. We apply this theory to prove the junta approximation theorem in the fourth subsection. Then in the final subsection, we complete our proof via a bootstrapping argument based on Shearer's entropy inequality.

7.1 | Hypergraphs

When m is huge, it is natural to view a code $\mathcal{F} \subset [m]^n$ as an n-graph (n-uniform hypergraph) which is n-partite (each edge has one vertex in each part) with parts $V_1, \dots, V_n$, where each $V_i = \{(i,a) \mid a \in [m]\}$, identifying any $x \in [m]^n$ with $\{(i,a) : x_i = a\}$. This setting is most convenient for introducing general agreement configurations in the following definition, as these are a natural partite variation on the well-studied topic of expanded hypergraphs (see the survey [26]).

Definition 7.1. An ℓ-configuration is a pair $(\mathcal{H}, \mathcal{P})$ where $\mathcal{H}$ is a multi-ℓ-graph and $\mathcal{P} = (U_1, \dots, U_\ell)$ is a partition of $V(\mathcal{H})$ such that each edge has one vertex in each part. We identify any $\mathcal{H}$ with its multiset of edges $\{e_1, \dots, e_h\}$, so its size $h = |\mathcal{H}|$ is its number of edges. We often omit $\mathcal{P}$ from our notation. The density of $\mathcal{H}$ (with respect to $\mathcal{P}$) is $\mu(\mathcal{H}) = |\mathcal{H}| \prod_{i \in \ell} |U_i|^{-1}$. The kernel of $\mathcal{H}$ is $K(\mathcal{H}) = \bigcap_{i=1}^h e_i$.

The n-expansion $\mathcal{H}^+(n)$ of $\mathcal{H}$ is the n-configuration obtained by adding disjoint sets S_j of $n - \ell$ new vertices to each e_j, forming new parts $U_{\ell+1}, \dots, U_n$ so that each S_j has one vertex in each new part. We say that $\mathcal{F}_1, \dots, \mathcal{F}_h \subset [m]^n$ cross contain $\mathcal{H}$ if they do so for $\mathcal{H}^+(n)$ when viewed as n-graphs, that is, there are $x^j \in \mathcal{F}_j$ for $j \in [h]$ and an injection $\Phi : [\ell] \to [n]$, so that for any $j, j' \in [h]$ and $i \in [n]$, we have $x_i^j = x_i^{j'}$ exactly when $i = \Phi(k)$ for some $k \in [\ell]$ and $e_j \cap e_{j'} \cap U_k \neq \emptyset$. We say that $x_1, \dots, x_h$ realise $\mathcal{H}$ in $\mathcal{F}_1, \dots, \mathcal{F}_h$. If $\mathcal{F}_i = \mathcal{F}$ for all i, we say that $\mathcal{F}$ contains $\mathcal{H}$; otherwise, we say that $\mathcal{F}$ is $\mathcal{H}$-free.

Example 7.2. A code $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding if when viewed as an n-partite n-graph, it does not contain two edges e, e' with $|e \cap e'| = t-1$; equivalently, $\mathcal{F}$ is $\mathcal{H}$-free where $\mathcal{H}$ is the multi-$(t-1)$-graph with two identical edges.

The main result of this subsection is the following extremal result for cross containment at constant densities (this suffices for our purposes, so we do not investigate the optimal bound).

Lemma 7.3. *For any $\ell, h \in \mathbb{N}$, there is $C > 0$ so that if $\mathcal{H}$ is an ℓ-configuration of size h and $\mathcal{F}_1, \dots, \mathcal{F}_h \subset [m]^n$ with each $\mu(\mathcal{F}_i) > \varepsilon$, where $n > C \log(\varepsilon^{-1})$ and $m > 2hn/\varepsilon$, then $\mathcal{F}_1, \dots, \mathcal{F}_h$ cross contain $\mathcal{H}$.*

The proof will reduce to the case when $\mathcal{H}$ is a matching, as in the following claim.

Claim 7.4. If $\mathcal{F}_1, \dots, \mathcal{F}_h \subset [m]^n$ with $m > hn/\varepsilon$ and each $\mu(\mathcal{F}_i) > \varepsilon$ then $\mathcal{F}_1, \dots, \mathcal{F}_h$ cross contain a matching.

Proof. We choose disjoint edges $e_i \in \mathcal{F}_i$ for $i \geqslant 1$ according to a greedy algorithm. Each choice reduces the density of any $\mathcal{F}_i$ by at most $n/m < \varepsilon/h$, so the algorithm can be completed. □

Proof of Lemma 7.3. Write $\mathcal{H} = \{e_1, \dots, e_h\}$ and let $(U_1, \dots, U_\ell)$ be the fixed partition of $\mathcal{H}$. We identify each $\mathcal{F}_i$ with an n-partite n-graph with parts $V_i = \{(i, a) : a \in [m]\}$. We consider uniformly random injections $\Phi : [\ell] \to [n]$ and $\phi_j : U_j \to V_{\Phi(j)}$ for each $j \in [\ell]$. Each edge e_i then defines a restriction $\mathcal{G}_i = (\mathcal{F}_i)_{\Phi([\ell]) \to \alpha^i}$, where $\alpha^i_{\Phi(j)} = \phi_j(e_i \cap U_j)$ for $j \in [\ell]$.

We let $C = C(\ell, 1/2h)$ be as in Proposition 6.1, which is then applicable as $\mu(\mathcal{F}_i) \geqslant \varepsilon \geqslant e^{-n/C}$, giving $\mathbb{P}[\mu(\mathcal{G}_i) \geqslant (1 - 1/2h)\mu(\mathcal{F}_i)] \geqslant 1 - 1/2h$. By a union bound, we can fix Φ and $\phi_1, \dots, \phi_j$ so that all $\mu(\mathcal{G}_i) > \varepsilon/2$. Then $\mathcal{G}_1, \dots, \mathcal{G}_h$ cross contain a matching, so $\mathcal{F}_1, \dots, \mathcal{F}_h$ cross contain $\mathcal{H}$. □

7.2 | Shadows

In this subsection, we define our shadow (projection) operation and establish its two key properties mentioned above (boosting measure and strengthening uncapturability).

Definition 7.5. For $\mathcal{F} \subset [m]^n$ and $i \in [n]$, the i-shadow of $\mathcal{F}$ is $\partial_i(\mathcal{F}) = \bigcup_{a \in [m]} \partial_{i \to a}(\mathcal{F})$, where $\partial_{i \to a}(\mathcal{F}) = \mathcal{F}_{i \to a} \subset [m]^{n-1}$. For $I \subset [n]$, we let ∂_I be the composition (in any order) of $(\partial_i : i \in I)$.

The next lemma, analogous to a lemma for hypergraphs in [20], shows that shadows have significantly larger measure on average if we forbid a configuration with the following 'flatness' property.

Definition 7.6. The centre of a configuration is the set of vertices contained in more than one edge.

We say that a configuration is flat if each part has at most one vertex in the centre.

Lemma 7.7. *Suppose that $\mathcal{H}$ is a flat ℓ-configuration of size h and $\mathcal{F} \subset [m]^n$ is $\mathcal{H}$-free, with $n \geqslant h\ell$. Then $|\mathcal{F}| \leqslant h \sum_{i=1}^n |\partial_i(\mathcal{F})|$.*

Proof. Let $\mathcal{F}'$ be obtained from $\mathcal{F}$ by the following iterative deletion procedure starting from $\mathcal{F}' = \mathcal{F}$: if there is any $i \in [n]$ and $y \in \partial_i(\mathcal{F}')$ such that at most h choices of $x \in \mathcal{F}'$ with $x_{[n]\setminus i} = y$, then we delete all such x. Any $y \in \partial_i(\mathcal{F})$ is considered at most once in this procedure before it is removed from the shadow. Thus, the number of deleted sets is at most $h \sum_{i=1}^n |\partial_i(\mathcal{F})|$, so it suffices to show $\mathcal{F}' = \emptyset$.

Suppose for contradiction $\mathcal{F}' \neq \emptyset$. We will show that $\mathcal{F}$ contains $\mathcal{H}$. We write $\mathcal{H} = \{e_1, \dots, e_h\}$, denote the parts of $\mathcal{H}$ by $U_1, \dots, U_\ell$ and fix $u_j \in U_j$ for each $j \in [\ell]$ so that each vertex of U_j other than u_j is contained in at most one edge. Fix any $x \in \mathcal{F}'$. We will construct $x^1, \dots, x^h \in \mathcal{F}'$ realising $\mathcal{H}$ according to injections $\phi_j : U_j \to [m]$ so that $x^i_j = \phi_j(e_i \cap U_j)$ and $x_j = \phi_j(u_j)$ for all $i \in [h]$ and $j \in [\ell]$. As $\mathcal{H}$ is flat, this can be achieved greedily. Indeed, to construct x^i, we can start from $x^i = x$ and one by one for each j such that $e_i \cap U_j \neq \{u_j\}$ replace x^i_j by some new value not yet used in coordinate j, which is possible as there are at least $h + 1$ choices for x^i_j for any given $x^i_{[n]\setminus\{j\}}$. However, $\mathcal{F}$ is $\mathcal{H}$-free, so we have the required contradiction. □

We conclude this subsection by showing under the same conditions as the previous lemma, that if a code is uncapturable, then it has some shadow which is significantly more uncapturable. The key point is that the uncaptured measure is increased by a factor $\Omega(m/n)$, albeit at the expense of only considering restrictions that are n times smaller.

Lemma 7.8. *Suppose that $\mathcal{H}$ is a flat ℓ-configuration of size h and $\mathcal{F} \subset [m]^n$ is $\mathcal{H}$-free, with $n \geqslant h\ell$. If $\mathcal{F}$ is (r, ε)-uncapturable, then $\partial_i(\mathcal{F})$ is $(r/n, \varepsilon m/nh)$-uncapturable for some $i \in [n]$.*

Proof. We suppose that each $\partial_i(\mathcal{F})$ is $(r/n, \delta)$-capturable and show that $\delta \geqslant \varepsilon m/nh$. By definition, for each $i \in [n]$, there is a collection $\mathcal{D}_i$ of at most r/n dictators in $[m]^{[n]\setminus\{i\}}$ such that $\mu(\partial_i(\mathcal{F}) \setminus \bigcup \mathcal{D}_i) \leqslant \delta$. We let $\mathcal{D} = \bigcup_{i=1}^n \mathcal{D}_i$ where now we consider each dictator in $[m]^n$. Then $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \geqslant \varepsilon$ by uncapturability. Applying Lemma 7.7 to $\mathcal{F} \setminus \bigcup \mathcal{D}$, noting that each $\partial_i(\mathcal{F} \setminus \bigcup \mathcal{D}) \subset \partial_i(\mathcal{F}) \setminus \bigcup \mathcal{D}_i$, we have

$$|\mathcal{F} \setminus \bigcup \mathcal{D}| \leqslant s \sum_{i=1}^{n} |\partial_i(\mathcal{F}) \setminus \bigcup \mathcal{D}_i| \leqslant hn \cdot \delta m^{n-1},$$

so $\varepsilon \leqslant \mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant \delta hn/m$, that is, $\delta \geqslant \varepsilon m/nh$. □

7.3 | Kernels and iterated shadows

In this subsection, we consider configurations with a non-trivial kernel (intersection of all edges), for which we show that they remain free of some configuration under iterated shadows (as many as the size of the kernel), so the results of the previous subsection on single shadows become correspondingly stronger in this setting. Firstly we introduce some convenient notation.

Definition 7.9. Given an ℓ-configuration $\mathcal{H}$, we write $\mathcal{H} \oplus [t]$ for the $(\ell + t)$-configuration with t additional parts of size 1 where each edge of $\mathcal{H}$ is extended to also include the t new vertices.

Given an ℓ-configuration $\mathcal{H}$ on v vertices, we let $\mathsf{flat}(\mathcal{H})$ be the v-configuration obtained by taking a copy of $\mathcal{H}$ with one vertex in each part and adding to each edge e disjoint sets S_e of $v - \ell$ new vertices.

As the notation suggests, $\mathsf{flat}(\mathcal{H})$ is flat, as each of the original vertices gets its own part, and each new vertex is in exactly one edge. To illustrate the construction, consider the 2-configuration $\mathcal{H}$ that is a 4-cycle on [4] with parts $\{1, 2\}$ and $\{3, 4\}$ and edges $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$ and $\{2, 4\}$. Every vertex of $\mathcal{H}$ is in two edges, so in the centre, and so $\mathcal{H}$ is not flat. To construct $\mathsf{flat}(\mathcal{H})$, which is a 4-configuration, we add two new vertices to each edge $\{i, j\} \in \mathcal{H}$, which we call x_{ij}^k for $k \in [4] \setminus \{i, j\}$, and form parts U_k for $k \in [4]$ where each U_k consists of k and the two vertices of the form x_{ij}^k.

Remark 7.10.

(1) Any (flat) configuration with a kernel of size t may be expressed as $\mathcal{H} \oplus [t]$ for some (flat) configuration $\mathcal{H}$ with no kernel.
(2) If $\mathcal{H}$ is flat and contained in $\mathcal{H}'$, then $\mathcal{H}$ is contained in $\mathsf{flat}(\mathcal{H}')$.

Lemma 7.11. *For any (flat) configuration $\mathcal{H}$, there exists a (flat) configuration $\mathcal{H}'$ such that for all $t \in \mathbb{N}$, there exist $m_0, n_0 \in \mathbb{N}$ such that if $\mathcal{F} \subset [m]^n$ with $n \geqslant n_0$, $m \geqslant m_0$ is $\mathcal{H} \oplus [t]$-free, then $\partial_i(\mathcal{F})$ is $\mathcal{H}' \oplus [t-1]$-free for all $i \in [n]$.*

Proof. Consider any ℓ-configuration $\mathcal{H}$ of size h. We let $C = C(\ell + t, h)$ be as in Lemma 7.3, $\varepsilon = 1/2h$, $n_1 = 2C \log(\varepsilon^{-1})$, $m_1 = 3hn/\varepsilon$, and then prove the statement for $\mathcal{H}' = [m_1]^{n_1}$, the complete n_1-partite n_1-graph with parts of size m_1. We note that if $\mathcal{H}$ was flat initially, then one can take

$\mathsf{flat}(\mathcal{H}')$ instead of $\mathcal{H}'$ to preserve flatness, and the correctness follows from the analysis below and by Remark 7.10.

We show the contrapositive statement, that is, that if $\partial_{i^*}(\mathcal{F})$ contains $\mathcal{H}' \oplus [t-1]$ for some $i^* \in [n]$, then $\mathcal{F}$ contains $\mathcal{H} \oplus [t]$. The version for flat configurations will then follow by Remark 7.10.2.

By relabelling, we can assume that we have $X = \{x(y) : y \in [m_1]^{n_1}\} \subset \mathcal{F}$ where each $x(y)_{[n_1]} = y$, there is some $T \in \binom{[n]\setminus[n_1]}{t-1}$ such that $x(y)_i = 1$ for all $i \in T$, and $x(y)_i \neq x(y')_i$ whenever $y \neq y'$ and $i \notin [n] \setminus (T \cup [n_1] \cup \{i^*\})$. We m-colour $[m_1]^{n_1}$ as $C_1, \dots, C_m$, where each $C_j = \{y : x(y)_{i^*} = j\}$.

Note that if some $\mu(C_j) \geqslant \varepsilon$, then applying Lemma 7.3 with $\mathcal{F}_1 = \cdots = \mathcal{F}_h = C_j$, we find a copy of $\mathcal{H}$ in C_j. The corresponding $x(y) \in \mathcal{F}$ for each y in this copy agrees outside $[n_1]$ in coordinates $T \cup \{i^*\}$ and no others, so we obtain a copy of $\mathcal{H} \oplus [t]$ in $\mathcal{F}$.

Thus, we may assume that each $\mu(C_j) < \varepsilon$. By repeated merging, we can form 'meta-colours' $C'_1, \dots, C'_{m'}$, each of which is a union of some of the C_j's, such that each $\mu(C'_j) \in (\varepsilon, 2\varepsilon)$, so $m' \geqslant 1/2\varepsilon = h$. By Lemma 7.3, $C'_1, \dots, C'_h$ cross contain $\mathcal{H} \oplus [1]$. The corresponding $x(y) \in \mathcal{F}$ for each y in this copy agree outside $[n_1]$ in coordinates T and no others, so again we obtain a copy of $\mathcal{H} \oplus [t]$ in $\mathcal{F}$. □

The following corollary is immediate by iterating Lemma 7.11.

Corollary 7.12. *For any (flat) configuration $\mathcal{H}$ and $t \in \mathbb{N}$, there exist $m_0, n_0 \in \mathbb{N}$ and a (flat) configuration $\mathcal{H}'$ such that if $\mathcal{F} \subset [m]^n$ with $n \geqslant n_0$, $m \geqslant m_0$ is $\mathcal{H} \oplus [t]$-free, then $\partial_I(\mathcal{F})$ is $\mathcal{H}'$-free for all $I \in \binom{[n]}{t}$.*

We also have the following corollary giving improved estimates on measures and uncapturability of iterated shadows.

Corollary 7.13. *For any flat configuration $\mathcal{H}$ and $t \in \mathbb{N}$, there exist $m_0, n_0 \in \mathbb{N}$ and $C > 0$ such that for any $\mathcal{H} \oplus [t-1]$-free $\mathcal{F} \subset [m]^n$ with $n \geqslant n_0$, $m \geqslant m_0$,*

(1) $|\mathcal{F}| \leqslant C \sum_{I \in \binom{[n]}{t}} |\partial_I(\mathcal{F})|$,

(2) *if $\mathcal{F}$ is (r, ε)-uncapturable, then $\partial_I(\mathcal{F})$ is $(r/n^t, (m/n)^t \varepsilon / C)$-uncapturable for some $I \in \binom{[n]}{t}$.*

Proof. We argue by induction on t. The base case $t = 1$ is given by Lemmas 7.7 and 7.8. Now suppose $t \geqslant 2$. By Lemma 7.11, there is a configuration $\mathcal{H}'$ depending only on $\mathcal{H}$ such that each $\partial_i(\mathcal{F})$ is $\mathcal{H} \oplus [t-2]$-free. The induction hypothesis of (1) gives $C' = C(\mathcal{H}', t-1)$ such that each $|\partial_i(\mathcal{F})| \leqslant C' \sum\{|\partial_{I \cup \{i\}}(\mathcal{F})| : I \in \binom{[n]}{t-1}\}$, which proves (1). For (2), if $\mathcal{F}$ is (r, ε)-uncapturable, then by Lemma 7.8, the family $\partial_i(\mathcal{F})$ is $(r/n, \varepsilon m/n|\mathcal{H}|)$-uncapturable for some $i \in [n]$. By the induction hypothesis of (2), $\partial_{I \cup \{i\}}(\mathcal{F})$ is $(r/n^t, (m/n)^t \varepsilon / C)$-uncapturable for some $I \in \binom{[n]\setminus\{i\}}{t-1}$, which proves (2). □

7.4 | Junta approximation

In this subsection, we prove Theorem 1.2 in the case that m is huge (at least exponential in n).

Theorem 7.14. *For any $t, N \in \mathbb{N}$, there are $K, n_0 \in \mathbb{N}$ such that if $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding with $n \geqslant n_0$ and $m \geqslant 2^{n/N}$, then there exists a subcube D of co-dimension t such that $\mu(\mathcal{F} \setminus D) \leqslant 2^{-n/K} m^{-t}$.*

Proof. We apply Lemma 5.6 with $r = n^t$, $k = t$ and $\varepsilon = 2^{-2n/K} \geqslant 1/m$, where $K, n_0 \gg t, N$, obtaining a collection $\mathcal{D}$ of at most $r^k = n^{t^2}$ subcubes of co-dimension at most t such that $\mathcal{F}_{R\to\alpha}$ is $(r, \varepsilon\mu(D)^{-1}m^{-t})$-uncapturable for each $D = D_{R\to\alpha} \in \mathcal{D}$ and $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant n^{2t^2}\varepsilon m^{-t}$. We let $\mathcal{D}_d$ be the set of subcubes in $\mathcal{D}$ of co-dimension d. To prove the theorem, it suffices to show that (a) $\mathcal{D}_d = \emptyset$ for $d < t$ and (b) $|\mathcal{D}_t| \leqslant 1$.

To see (a), suppose for a contradiction that $D_{R\to\alpha} \in \mathcal{D}_{t-1-s}$ with $s \geqslant 0$. As $\mathcal{F}$ is $(t-1)$-avoiding, $\mathcal{F}_{R\to\alpha}$ is s-avoiding, that is, is $\mathcal{H} \oplus [s]$-free, where $\mathcal{H}$ is the flat 0-configuration consisting of two copies of the empty set. By Corollary 7.12, there is some flat configuration $\mathcal{H}'$ such that $\partial_I(\mathcal{F}_{R\to\alpha})$ is $\mathcal{H}'$-free for any $I \in \binom{[n]\setminus R}{s}$. By Corollary 7.13, as $\mathcal{F}_{R\to\alpha}$ is $(n^t, \varepsilon m^{-s-1})$-uncapturable, there is some $I \in \binom{[n]\setminus R}{s}$ such that $\mathcal{G} := \partial_I(\mathcal{F}_{R\to\alpha})$ is $(n^{t-s}, \varepsilon/O_t(mn^s))$-uncapturable.

To obtain the required contradiction, we will show that $\mathcal{G}$ contains $\mathcal{H}'$. Write $|\mathcal{H}'| = h' = O_t(1)$. Let $\mathcal{J}$ be the set of all dictators $D_{i\to a}$ such that $\mu(\mathcal{G}_{i\to a}) > \varepsilon^2/n^2$. We claim that $|\mathcal{J}| < h'$. To see this, suppose on the contrary that $\mathcal{J}$ contains $D_{i^1\to a^1}, \ldots, D_{i^{h'}\to a^{h'}}$. Let $I' = \{i^1, \ldots, i^{h'}\}$. By averaging, we can fix $x^j \in [m]^{I'}$ for $j \in [h']$ such that $x^j_{i^j} = a^j$, $x^j_{i^{j'}} \neq a^{j'}$ for all $j' \neq j$, so that $\mathcal{F}_j = \mathcal{G}_{I'\to x^j}$ has $\mu(\mathcal{F}_j) > \varepsilon^2/2n^2$. However, then $\mathcal{H}'$ is cross contained in $\mathcal{F}_1, \ldots, \mathcal{F}_{h'}$ by Lemma 7.3, applied with $\varepsilon^2/2n^2$ in place of ε (using $n > C\log(2n^2/\varepsilon^2)$ and $m > 2hn \cdot 2n^2/\varepsilon^2$ for large K). Thus, $|\mathcal{J}| < h'$, as claimed.

By uncapturability of $\mathcal{G}$, writing $\mathcal{G}' = \mathcal{G} \setminus \bigcup \mathcal{J}$, we have $\mu(\mathcal{G}') \geqslant \varepsilon/O_t(mn^s) > \varepsilon^2/m$. By Lemma 7.7, we can fix $i^* \in [n] \setminus (R \cup I)$ with $|\mathcal{G}'|/h'n \leqslant |\partial_{i^*}(\mathcal{G}')|$. We fix any partition $(\mathcal{F}'_a : a \in [m])$ of $\partial_{i^*}(\mathcal{G}')$ such that each $\mathcal{F}'_a \subset \partial_{i^*\to a}(\mathcal{G}')$. Then $\sum_a \mu(\mathcal{F}'_a) = \mu(\partial_{i^*}\mathcal{G}') \geqslant \mu(\mathcal{G}')m/h'n > \varepsilon^2/h'n$. Also, by definition of $\mathcal{J}$, each $\mu(\mathcal{F}'_a) < \varepsilon^2/n^2$. By repeated merging, we can form a partition $\mathcal{P}$ of $[m]$ such that each $S \in \mathcal{P}$ has $\sum_{a\in S} \mu(\mathcal{F}'_a) \in (\varepsilon^2/n^2, 2\varepsilon^2/n^2)$. Then $|\mathcal{P}| \geqslant h'$, so we can choose $S_1, \ldots, S_{h'}$ in $\mathcal{P}$ and apply Lemma 7.3 to see that $\mathcal{F}_1, \ldots, \mathcal{F}_{h'}$ cross contain $\mathcal{H}'$, where each $\mathcal{F}_i = \bigcup_{a\in S_i} \mathcal{F}'_a$. This completes the proof of (a).

To see (b), suppose for contradiction that we have distinct subcubes $D_{R_j\to\alpha_j}$ for $j = 1, 2$ of co-dimension t. Suppose that they agree on $t-1-s$ coordinates, for some $s \geqslant 0$. Consider $\mathcal{G}_1 = \mathcal{F}_{R_1\to\alpha_1} \setminus \mathcal{F}_{R_2\to\alpha_2}$ and $\mathcal{G}_2 = \mathcal{F}_{R_2\to\alpha_2} \setminus \mathcal{F}_{R_1\to\alpha_1}$. By uncapturability, both $\mu(\mathcal{G}_j) \geqslant \varepsilon > e^{-n/C}$, where $C = C(s, 0.1)$ is as in Proposition 6.1, as K is large. Consider uniformly random $\mathbf{S} \sim \binom{[n]\setminus(R_1\cup R_2)}{s}$ and $\mathbf{x} \in [m]^{\mathbf{S}}$.

By Proposition 6.1, both $\mathbb{P}[\mu((\mathcal{G}_j)_{\mathbf{S}\to\mathbf{x}}) \geqslant 0.9\mu(\mathcal{G}_j)] \geqslant 0.9 - o(1)$, as $\mathbf{S}$ is total variation distance $o(1)$ from uniform on $\binom{[n]\setminus R_j}{s}$. Thus, we can fix S, x so that both $\mathcal{G}'_j = (\mathcal{G}_j)_{S\to x}$ have $\mu(\mathcal{G}'_j) > 0.9\varepsilon$. By averaging, we can fix some $\mathcal{G}''_1 = (\mathcal{G}'_1)_{R_2\setminus R_1\to a^1}$ with $\mu(\mathcal{G}''_1) \geqslant \mu(\mathcal{G}'_1) > 0.9\varepsilon$, and similarly some $\mathcal{G}''_2$. Then $\mathcal{G}''_1$, $\mathcal{G}''_2$ are defined by restrictions of $\mathcal{F}$ to $R_1 \cup R_2 \cup S$ with agreement exactly $t-1$, so must be cross intersecting. However, $m \geqslant 2^{n/N} \gg \varepsilon^{-1}$ for large K, so this contradicts Lemma 4.6. $\square$

7.5 | Bootstrapping

We conclude this part with the bootstrapping step that completes the proof of our main theorem for huge alphabets, which we restate as follows.

Theorem 7.15. *For any $t, N \in \mathbb{N}$, there is $n_0 \in \mathbb{N}$ such that if $n \geqslant n_0$, $m \geqslant 2^{n/N}$ and $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding, then $|\mathcal{F}| \leqslant m^{n-t}$, with equality only when $\mathcal{F}$ is a subcube of co-dimension t.*

We require Shearer's entropy lemma [3], as applied to the projection operators $\Pi_S = \partial_{[n]\setminus S}$ on $[m]^n$.

Lemma 7.16. *For $\mathcal{F} \subset [m]^n$ and $k \in [n]$, we have $|\mathcal{F}|^{\binom{n-1}{k-1}} \leqslant \prod_{|S|=k} |\Pi_S(\mathcal{F})|$.*

Proof of Theorem 7.15. Suppose $\mathcal{F} \subset [m]^n$ is $(t-1)$-avoiding with $|\mathcal{F}| \geqslant m^{n-t}$. By Theorem 7.14, there is a subcube D of co-dimension t such that $\mathcal{G} := \mathcal{F} \setminus D$ has $\varepsilon := \mu(\mathcal{G})m^t \leqslant 2^{-n/K}$, for some $K = K(N,t)$. We may assume $D = \{x \in [m]^n \mid x_1 = \cdots = x_t = 1\}$. Suppose for contradiction that $\varepsilon > 0$. For each $T \subsetneq [t]$, let $\mathcal{G}_T$ be the set of all $x_{[n]\setminus[t]}$ where $x \in \mathcal{G}$ with $T = \{i \in [t] : x_i = 1\}$. We have $\varepsilon = \mu(\mathcal{G})m^t \leqslant \sum_T m^{t-|T|}\mu(\mathcal{G}_T)$, so for a contradiction, it suffices to show that each $\mu(\mathcal{G}_T) < m^{|T|-t}\varepsilon/n$.

As $\mathcal{F}$ is $(t-1)$-avoiding, each $\mathcal{G}_T$ is $(t-1-|T|)$-avoiding. In particular, if $|T| = t-1$, then $\mathcal{G}_T$ is intersecting, so by Lemma 4.6, we have the required bound $\mu(\mathcal{G}_T) < 2\varepsilon/m^2 < m^{-1}\varepsilon/n$.

Now fix any $T \subset [t]$ where $|T| = t-1-d$ with $d \geqslant 1$. As $\mathcal{G}_T$ is d-avoiding, it is free of a configuration with kernel size d, so by Corollary 7.13, we have $|\mathcal{G}_T| \leqslant O_t(1)\sum_{I\in\binom{[n]}{d+1}} |\partial_I(\mathcal{G}_T)|$. Fix any $I \in \binom{[n]}{d+1}$. To complete the proof, it suffices to establish the following claim, as this will imply $\mu(\mathcal{G}_T) < O_t(1)(n/m)^{d+1}\varepsilon^2 < m^{|T|-t}\varepsilon/n$.

Claim 7.17. $\mu(\partial_I\mathcal{G}_T) < \varepsilon^2$.

We will prove this claim using Shearer's inequality with $k = d$, so we now analyse the projections $\Pi_S\partial_I\mathcal{G}_T = \Pi_S\mathcal{G}_T$ for $S \in \binom{[n]\setminus I}{d}$. For such S with $S \cap [t] \neq \emptyset$, we use the trivial bound $|\Pi_S\mathcal{G}_T| \leqslant m^d$. Now fix S with $S \cap [t] = \emptyset$. We will show that $\mu(\Pi_S\mathcal{G}_T) < 2\varepsilon$.

To see this, we first show for any $x \in \Pi_S\mathcal{G}_T$ that $\mathcal{F}'_x := \mathcal{F}_{[t]\to\mathbf{1},S\to x}$ has $\mu(\mathcal{F}'_x) \leqslant n/m$. Suppose not, and fix $y \in \mathcal{G}$ with $\pi_S(y) = x$ and $T = \{i \in [t] : y_i = 1\}$. By a union bound $\mu(\mathcal{F}'_x \setminus \bigcup_{i\in[n]\setminus([t]\cup S)} D_{i\to y_i}) > 0$, so we can choose $z \in \mathcal{F}'_x$ that disagrees with y on $[n] \setminus ([t] \cup S)$. However, extending z with $x \in [m]^S$ and $\mathbf{1} \in [m]^t$ gives $z^+ \in \mathcal{F}$ with $\mathrm{agr}(z^+, y) = t-1$, which is impossible, so indeed $\mu(\mathcal{F}'_x) \leqslant n/m$.

As $|\mathcal{F}| \geqslant |D_{[t]\to\mathbf{1}}|$, this implies $|\mathcal{G}| \geqslant |D_{[t]\to\mathbf{1}} \setminus \mathcal{F}| \geqslant |\Pi_S\mathcal{G}_T| \cdot (1-n/m)m^{n-t-d}$, so $\varepsilon = \mu(\mathcal{G})m^t \geqslant (1-n/m)\mu(\Pi_S\mathcal{G}_T)$, giving $\mu(\Pi_S\mathcal{G}_T) < 2\varepsilon$. Finally, writing $n' = |[n] \setminus I| = n-(d+1)$, Lemma 7.16 gives

$$|\partial_I\mathcal{G}_T|^{\binom{n'-1}{d-1}} \leqslant \prod_S |\Pi_S\partial_I\mathcal{G}_T)| \leqslant (m^d)^{\binom{n'}{d}-\binom{n'-t}{d}}(2\varepsilon m^d)^{\binom{n'-t}{d}} = (2\varepsilon)^{\binom{n'-t}{d}}(m^d)^{\binom{n'}{d}},$$

so $|\partial_I\mathcal{G}_T| \leqslant (2\varepsilon)^{n'/2d}(m^d)^{n'/d} < \varepsilon^2 m^{n'}$. This completes the proof of the claim, and so of the theorem. □

Part IV: Generalisations and open problems

8 | CONFIGURATIONS

In this section, we briefly consider generalisations to excluded configurations (as in the previous section). Our aim is not a systematic study, but just to illustrate the further potential applications of our methods. We start with a general junta approximation result for small alphabets.

Theorem 8.1. *For every $\eta > 0$, configuration $\mathcal{H}$ and $m \in \mathbb{N}$ with $m > |\mathcal{H}|$, there are $J, n_0 \in \mathbb{N}$ such that if $\mathcal{F} \subset [m]^n$ is $\mathcal{H}$-free with $n \geqslant n_0$, then there is an $\mathcal{H}$-free J-junta $\mathcal{J} \subset [m]^n$ such that $\mu(\mathcal{F} \setminus \mathcal{J}) \leqslant \eta$.*

The proof requires the following generalisation of Theorem 3.4.

Theorem 8.2. *For every $h, m \in \mathbb{N}$ with $m > h$ and $\mu > 0$, there are $\varepsilon, c > 0$ and $r \in \mathbb{N}$ such that if $\mathcal{F}_1, \dots, \mathcal{F}_h \subset [m]^n$ are (r, ε)-pseudorandom with each $\mu(\mathcal{F}_j) > \mu$ and $(x_1, \dots, x_h) \in ([m]^n)^h$ is uniformly random subject to $\mathsf{agr}(x_j, x_{j'}) = 0$ whenever $j \neq j'$ then $\mathbb{P}(x_1 \in \mathcal{F}_1, \dots, x_h \in \mathcal{F}_h) > c$.*

The proof of Theorem 8.2 is the same as that of Theorem 3.4, except that the absolute spectral gap condition must be replaced by a more general condition on 'correlated spaces', which specialises to our situation as follows. Given $f, g : [m]^h \to \mathbb{R}$ with $\mathbb{E}f = \mathbb{E}g = 0$ and $\mathbb{E}f^2 = \mathbb{E}g^2 = 1$, such that f depends only on the first coordinate and g does not depend on the first coordinate, and $\mathbf{a} \in [m]^h$ with distinct coordinates chosen uniformly at random, we need to show that $\mathbb{E}f(\mathbf{a})g(\mathbf{a}) < 1$. By considering the equality conditions for Cauchy–Schwarz, it is not hard to see that this holds when $m > h$.

Proof of Theorem 8.1. The proof is the same as that of Theorem 2.1, except that instead of showing that $\mathcal{J}$ is t-intersecting, we need to show that $\mathcal{J}$ is $\mathcal{H}$-free. To see this, we suppose for a contradiction that $\mathcal{J}$ contains $\mathcal{H}$ and show that $\mathcal{F}$ contains $\mathcal{H}$. We suppose $\mathcal{H} = \{e_1, \dots, e_h\}$ is an ℓ-configuration with parts $(U_1, \dots, U_\ell)$ realised by $x^1, \dots, x^h \in \mathcal{J}$. By relabelling, we can assume that $\mathcal{H}$ is realised on coordinate set $[\ell]$, that is, for any $j, j' \in [h]$ and $i \in [n]$, we have $x_i^j = x_i^{j'}$ exactly when $i \in [\ell]$ and $e_j \cap e_{j'} \cap U_i \neq \emptyset$. For $j \in [h]$, we let $\mathcal{G}_j = \mathcal{F}_{J \cup [\ell] \to x^j_{J \cup [\ell]}}$. Then each $\mathcal{G}_j$ is $(r - \ell, \varepsilon)$-pseudorandom with density at least $\eta/3$, so by Theorem 8.2, we find $w_j \in \mathcal{G}_j$ for $j \in [h]$ with $\mathsf{agr}(w_j, w_{j'}) = 0$ whenever $j \neq j'$. However, $((x^j_{J \cup [\ell]}, w_j) : j \in [h])$ realise $\mathcal{H}$ in $\mathcal{F}$, contradiction. $\square$

Next we will turn to large alphabets, for which we require the following generalised Hoffman bound.

Lemma 8.3. *Let $m > hb$ and suppose that ν is a b-balanced product measure on $[m]^n$ and $\mathcal{F}_1, \dots, \mathcal{F}_h \subset [m]^n$ with $\prod_{j=1}^h \nu(\mathcal{F}_j) > 2^h b/(m - hb) > 0$. Then $\mathcal{F}_1, \dots, \mathcal{F}_h$ cross contain an h-matching.*

Proof. We show the following statement by induction on h: if $(x^1, \dots, x^h) \in ([m]^n)^h$ is distributed as ν^h conditioned on $\mathsf{agr}(x^j, x^{j'}) = 0$ whenever $j \neq j'$, then $\mathbb{P}(x^1 \in \mathcal{F}_1, \dots, x^h \in \mathcal{F}_h) \geqslant \nu(\mathcal{F}_1) \dots \nu(\mathcal{F}_h) - 2^h b/(m - hb)$. The case $h = 1$ is trivial.

For the induction step, as in the proof of Lemma 5.9, we consider the product Markov chain T on $[m]^n$ where each T_i is the Markov chain on $[m]$ with transition probabilities $(T_i)_{xx} = 0$ and $(T_i)_{xy} = \nu_i(y)/(1 - \nu_i(x))$ for $y \neq x$. We also consider $y^1, \dots, y^{h-1}, x^h$, where x^h is chosen according to ν and each y^j is chosen independently according to ν conditioned on $\mathsf{agr}(y^j, x^h) = 0$. We write f_j for the characteristic functions of $\mathcal{F}_j$ for $j \in [h]$. We have

$$\mathbb{E}[f_1(y^1) \dots f_{h-1}(y^{h-1}) f_h(x^h)] = \mathbb{E}_x[Tf_1(x) \dots Tf_{h-1}(x) f_h(x)] = \mathbb{E}f_1 \dots \mathbb{E}f_h + \sum_{\emptyset \neq S \subset [h-1]} \mathbb{E}g_S,$$

where $g_S(x) = \prod_{i \in S}(Tf_i - \mathbb{E}f_i)(x) \prod_{i \in [h-1] \setminus S} f_i(x)$. For each such S, we fix some $s \in S$ and write $g_S(x) = (Tf_s - \mathbb{E}f_s)(x) h_S(x)$. As $Tf_s - \mathbb{E}f_s = T(f_s - \mathbb{E}f_s)$ and $\mathbb{E}(f_s - \mathbb{E}f_s) = 0$, as in the proof of Lemma 5.9, we have the spectral bound

$$\|Tf_s - \mathbb{E}f_s\|_2 \leqslant \frac{b/m}{1 - b/m} = \frac{b}{m-b}.$$

Then $|\mathbb{E}g_S(x)| \leqslant b/(m-b)$ by Cauchy–Schwarz, so $\mathbb{E}[f_1(y^1) \dots f_{h-1}(y^{h-1}) f_h(x^h)] \geqslant \mathbb{E}f_1 \dots \mathbb{E}f_h - 2^{h-1} b/(m-b)$.

Now we write $\mathbb{P}(x^1 \in \mathcal{F}_1, \dots, x^h \in \mathcal{F}_h) = \mathbb{E}\prod_{j=1}^h f_j(x^j) = \mathbb{E}_x f_h(x) \mathbb{E}[\prod_{j=1}^{h-1} f_j(x^j) \mid x^h = x]$. For each x, we apply the induction hypothesis to $f_1, \dots, f_{h-1}$ on $\{x \in [m]^n : \mathrm{agr}(x, x^h) = 0\}$, which is isomorphic to $[m-1]^n$, according to the product measure $\nu[x^h]$ with each $\nu[x^h]_i(a) = \nu_i(a)/(1 - \nu_i(x_i^h)) \leqslant \frac{b/m}{1-b/m} = \frac{b}{m-b}$, so $\nu[x^h]$ is b'-balanced, where $b' = b(m-1)/(m-b)$. As $b'/(m-1-(h-1)b') = b/(m-hb)$, by induction hypothesis $\mathbb{E}[\prod_{j=1}^{h-1} f_j(x^j) \mid x^h = x] \geqslant \mathbb{E}[\prod_{j=1}^{h-1} f_j(y^j) \mid x^h = x] - 2^{h-1} b/(m-hb)$, so $\mathbb{E}\prod_{j=1}^h f_j(x^j) \geqslant \mathbb{E}_x f_h(x) [\mathbb{E}[\prod_{j=1}^{h-1} f_j(y^j) \mid x^h = x] - 2^{h-1} b/(m-hb)] \geqslant \mathbb{E}f_1 \dots \mathbb{E}f_h - 2^h b/(m-hb)$. □

For moderate alphabets, we have the following generalised form of our earlier lemma on fixed agreements between uncapturable families: we show that uncapturable families cross contain any configuration.

Theorem 8.4. *For any configuration $\mathcal{H}$ of size h and $s, k \in \mathbb{N}$, there are $r, m_0, N \in \mathbb{N}$ such that if $m \geqslant m_0$, $n \geqslant N \log m$ and $\mathcal{A}_j \subset [m]^{[n] \setminus R_j}$ are (r, m^{-k})-uncapturable with $|R_j| \leqslant k$ for $j \in [h]$, then there is a realisation $y^1, \dots, y^h$ of $\mathcal{H}$ with $y^j = x^j_{[n] \setminus T}$ for some $x^j \in \mathcal{A}_j$ for $j \in [h]$, where $T = \bigcup_j R_j$.*

Proof. We follow the proof of Theorem 6.4.

Step 1: Globalness. We define $\mathcal{A}_j^t$ for $j \in [h]$, $0 \leqslant t \leqslant h$ as follows. Initially, all $\mathcal{A}_j^0 = \mathcal{A}_j$. At step $t \in [h]$, we apply Lemma 5.5 to $\mathcal{A}_t^{t-1}$, which will have $\mu(\mathcal{A}_t^{t-1}) \geqslant m^{-k}$, with $\gamma = m^{-1/10}$ and $r/100kh$ in place of r we obtain $\mathcal{A}_t^t = (\mathcal{A}_t^{t-1})_{R'_t \to \alpha'_t}$ that is $(r/100kh, \mu(\mathcal{A}_t^{t-1})/\gamma)$-global with $\mu(\mathcal{A}_t^t) \geqslant \mu(\mathcal{A}_t^{t-1})$, where $|R'_1| \leqslant (r/100kh) \log_{1/\gamma}(1/\mu(\mathcal{A}_t^{t-1})) \leqslant r/10h$. For each $j \in [h] \setminus \{t\}$, we let $\mathcal{A}_j^t = \mathcal{A}_j^{t-1} \setminus \bigcup_{i \in R'_t} D_{i \to \alpha'_t(i)}$. Then uncapturability implies the above assumption $\mu(\mathcal{A}_t^{t-1}) \geqslant m^{-k}$. By Claim 5.4, each $\mathcal{A}_t^t$ is $(\gamma m/4, \mu(\mathcal{A}_t^t)/2)$-uncapturable, so $\mu(\mathcal{A}_t^h) \geqslant \frac{1}{2}\mu(\mathcal{A}_t^t)$, which implies that $\mathcal{A}_t^h$ is $(r/100kh, 2\mu(\mathcal{A}_t^h)/\gamma)$-global.

Step 2: Fairness. As $n \geqslant N \log m$ and N is large, each $\mu(\mathcal{A}_j^h) \geqslant \frac{1}{2} m^{-k} \geqslant e^{-n/C}$, where $C = C(s, 1/2h)$ is as in Proposition 6.1. Consider uniformly random $\mathbf{L} \in \binom{[n] \setminus \bigcup_j (R_j \cup R'_j)}{\ell}$ and let $\mathbf{z}_1, \dots, \mathbf{z}_\ell \in [m]^{\mathbf{L}}$ be a uniformly random copy of $\mathcal{H}$. By Proposition 6.1, each $\mathbb{P}[\mu((\mathcal{A}_j^h)_{\mathbf{L} \to \mathbf{z}_j}) \geqslant \frac{1}{2}\mu(\mathcal{A}_j^h)] \geqslant 1 - 1/2h - o(1)$. Thus, we can fix L and $z_1, \dots, z_\ell$ so that all $C_j = (\mathcal{A}_j^h)_{L \to z_j}$ have $\mu(C_j) \geqslant \frac{1}{2}\mu(\mathcal{A}_j^h)$, so are $(r/100kh, 4\mu(C_j)/\gamma)$-global.

Step 3: Expansion. We define $C_j^t \subset [m_t]^n$ for $j \in [h]$, $0 \leqslant t \leqslant h$ as follows. Initially, all $C_j^0 = C_j$ and $m_0 = m$. At step $t \in [h]$, we apply Lemma 6.11 with $\varepsilon = 1/4kh$ and $b = b_t = 4^{2^t}$, to C_t^{t-1}, which will have $\mu^{\pi_{t-1}}(C_t^{t-1}) \geqslant \frac{1}{8} m^{-k}$, obtaining $\pi_t \in \Pi_{m_{t-1}, m_t, b_t}$ with $m_t = \Omega_k(m_{t-1})$, $\alpha''_t \in [m_t]^{R''_t}$, where $R''_t \subset [n] \setminus (R_t \cup R'_t \cup L)$ with $|R''_t| < O_k(\log m) \ll n$, such that $C_t^t := (C_t^{t-1})^{\pi_t}_{R''_t \to \alpha''_t}$ has $\mu^{\pi_t}(C_t^t) \geqslant 2m^{-1/2h}$. For each $j \in [h] \setminus \{t\}$, we let $C_j^t = \pi_t(C_j^{t-1}) \setminus \bigcup_{i \in R''_t} D_{i \to \alpha''_t}$.

For $j > t$, we can write $C_j^t = \pi_{\circ t}(\mathcal{X})$, where $\pi_{\circ t} = \pi_t \circ \dots \circ \pi_1$ and $\mathcal{X} = C_j \setminus \bigcup_{t' \leqslant t} \bigcup \{D_{i \to a} : i \in R''_{t'}, (\pi_i^{\circ t'}(a) = (\alpha''_{t'})_i\}$. As $\mu(C_j)$ is $(r/100kh, 4\mu(C_j)/\gamma)$-global, it is $(\gamma m/8, \mu(C_j)/2)$-uncapturable, so $\mu(\mathcal{X}) \geqslant \mu(C_j)/2 \geqslant \frac{1}{8}m^{-k}$. By Claim 6.8, this implies the above assumption $\mu^{\pi_{t-1}}(C_t^{t-1}) \geqslant \frac{1}{8}m^{-k}$. At the end of the process, each $\mu^{\pi_{\circ h}}(C_j^h) \geqslant \mu^{\pi_{\circ h}}(C_j^j) - O_{k,h}(m^{-1}\log m) \geqslant m^{-1/2h}$.

Step 4: Generalised Hoffman bound. By averaging, we can choose restrictions $\mathcal{G}_j \subset [m_1]^R$ of C_j^h for $j \in [h]$ where $R = L \cup \bigcup_j (R_j \cup R'_j \cup R''_j)$ such that all $\nu(\mathcal{G}_j) \geqslant m^{-1/2h}$, where $\nu = \mu^{\pi_{\circ h}}$ is b_h-balanced. By construction, the elements of $\mathcal{G}_j$ are of the form $\pi_{\circ h}(x^j_{[n]\setminus R})$ where $x^j \in \mathcal{A}_j$ form a copy of $\mathcal{H}$ on L and have no other agreements in $R \setminus \bigcup_j (R_j \cup R'_j \cup R''_j)$. As $\prod_j \nu(\mathcal{G}_j) \geqslant m^{-1/2} > 2^h b_h/(m_h - hb_h) > 0$, by Lemma 8.3, we can find a cross matching in $\mathcal{G}_1, \dots, \mathcal{G}_h$, which corresponds to $x^j \in \mathcal{A}_j$ such that $y^j = x^j_{[n]\setminus T}$ realise $\mathcal{H}$. $\square$

We conclude this section with a junta approximation result for configurations over large alphabets, where for simplicity, we restrict attention to flat configurations with no kernel. For this case, we obtain a result that is analogous to our junta approximation result in terms of 'crosscuts' of expanded hypergraphs in [16].

Firstly we give the appropriate definition of the crosscut for configurations. Let $\mathcal{H}$ be an ℓ-configuration of size h. The crosscut $\sigma(\mathcal{H})$ is the minimum number s such that there is a collection $\bigcup \mathcal{D}$ of s co-dimension 1 subcubes such that $\mathcal{H} \subseteq \bigcup \mathcal{D}$, among all collection $\mathcal{D}$ of s co-dimension 1 and each edge $e \in \mathcal{H}$ is contained in exactly one subcube in $\mathcal{D}$. Note that $\sigma(\mathcal{H}) > 1$ if and only if $\mathcal{H}$ has no kernel, that is, $K(\mathcal{H}) = \emptyset$.

Theorem 8.5. *For every $\eta > 0$ and flat configuration $\mathcal{H}$ with no kernel, there is C such that if $m, n > C$ and $\mathcal{F} \subset [m]^n$ is $\mathcal{H}$-free, then there is a collection $\mathcal{D}$ of fewer than $\sigma(\mathcal{H})$ subcubes of co-dimension 1 such that $\mu(\mathcal{F} \setminus \bigcup \mathcal{D}) \leqslant \eta/m$.*

Proof. Let $\mathcal{F} \subset [m]^n$ be $\mathcal{H}$-free, where $\mathcal{H} = \{e_1, \dots, e_h\}$ is an ℓ-configuration with parts $(U_1, \dots, U_\ell)$ and $K(\mathcal{H}) = \emptyset$.

Firstly we consider moderate alphabet sizes, that is, $n \geqslant N \log m$, with $m, N \gg h, \ell$. We can assume that $\mathcal{F}$ is (r, m^{-2})-capturable, with $h, \ell \ll r \ll m$; otherwise, we find $\mathcal{H}$ by Theorem 8.4, applied with all $\mathcal{A}_j = \mathcal{F}$. Thus, we find a collection $\mathcal{J}$ of at most r subcubes of co-dimension 1 such that $\mu(\mathcal{F} \setminus \bigcup \mathcal{J}) \leqslant m^{-2}$. Let $\mathcal{D}$ be the set of $D_{i\to a} \in \mathcal{J}$ such that $\mu(\mathcal{F}_{i \to a}) \geqslant \eta/2r$. Then $\mu(\bigcup \mathcal{J} \setminus \bigcup \mathcal{D}) < \eta/2m$, so it suffices to show $|\mathcal{D}| < \sigma(\mathcal{H})$.

Suppose for a contradiction that $|\mathcal{D}| \geqslant \sigma(\mathcal{H})$. Then by definition, $\mathcal{D}$ contains a copy of $\mathcal{H}$, without loss of generality realised on coordinate set $[\ell]$ by injections $\phi_i : U_i \to V_i = \{(i, a) : a \in [m]\}$, such that for each $j \in [h]$, there is $D_{i^j \to a^j} \in \mathcal{D}$ such that $\phi_{i^{j'}}(e_j \cap U_{i^{j'}}) = (i_{j'}, a^{j'})$ if and only if $j = j'$.

Let C be the set of $i \in [\ell]$ such that U_i contains a vertex in the centre of $\mathcal{H}$. For $i \in C$, let c_i be the vertex of U_i in the centre of $\mathcal{H}$ (which is unique by flatness). We may assume for any $i \in C$ that $\mathcal{D}$ either contains $D_{i \to \phi_i(c_i)}$ or does not contain any $D_{i \to a}$; indeed, if $\mathcal{D}$ does not contain $D_{i\to\phi_i(c_i)}$, then each $D_{i\to a}$ is $D_{i^j \to a^j}$ for at most one $j \in [h]$, so we can obtain an alternative realisation replacing ϕ_i by $\phi'_i : U_i \to V_{i'}$ for some new $i' \in [m]$, where $\phi'_i(v) = (i', a)$ whenever $\phi_i(v) = (i, a)$.

Let I be the set of all $i \in [n]$ such that $\mathcal{D}$ contains some $D_{i \to a}$. We claim that we can fix $y^j \in [m]^I$ for $j \in [h]$ such that (a) $\mu(\mathcal{F}_{I \to y^j}) \geqslant \eta/3r$ for all $j \in [h]$, (b) $y^j_{i^j} = a^j$ for all $j \in [h]$ and (c) for all $j \in [h]$, $j' \neq j$, if $i^{j'} = i^j$, then $y^j_i \neq y^{j'}_i$ for all $i \neq i^j$, and otherwise $y^j_i \neq y^{j'}_i$ for all $i \in I$ (in words,

y^j and $y^{j'}$ may only agree on i^j if the i's corresponding to j, j' coincide, and must disagree on any other coordinate). To see this, we apply a greedy algorithm. To define y^j, we consider

$$\mathcal{G}_j = \mathcal{F}_{i^j \to a^j} \setminus \bigcup_{\substack{j'<j, i\in I \\ (i, y_i^{j'}) \neq (i^j, a^j)}} \left(\left\{ x \in [m]^n \,|x_i = y_i^{j'} \right\}\right)_{i^j \to a^j},$$

which has $\mu(\mathcal{G}_j) \geqslant \mu(\mathcal{F}_{i^j \to a^j}) - \frac{h|I|}{m} \geqslant \frac{\eta}{2r} - \frac{h\ell}{m} > \frac{\eta}{3r}$. By averaging, we can fix a restriction $\mathcal{F}_{I \to y^j}$ of $\mathcal{G}_j$ with at least this measure, so the claim holds.

It remains to show that $\mathcal{G}_1, \dots, \mathcal{G}_h$ cross contain the configuration $\mathcal{H}'$ obtained from $\mathcal{H}$ by deleting the parts corresponding to I. As in the proof of Lemma 7.3, by Proposition 6.1, we can reduce to the case that $\mathcal{H}'$ is a matching, which holds by Lemma 8.3. Thus, the theorem holds for moderate alphabet sizes,

Now we consider huge alphabets, that is, $n \geqslant n_0$ and $m \geqslant 2^{n/N}$, where $K, n_0 \gg N$. We let $\mathcal{D}$ be the set of all dictators $D_{i \to a}$ such that $\mu(\mathcal{F}_{i \to a}) > \eta^2/n^2$. Similarly to the proof of (a) in Theorem 7.14, we have $|\mathcal{D}| < \sigma(\mathcal{H})$. Let $\mathcal{F}' = \mathcal{F} \setminus \bigcup \mathcal{D}$. It suffices to show $\mu(\mathcal{F}') < \eta/m$.

Suppose $\mu(\mathcal{F}') \geqslant \eta/m$. Similarly to the proof of (a) in Theorem 7.14, we fix $i^* \in [n]$ with $|\mathcal{F}'|/hn \leqslant |\partial_{i^*}(\mathcal{F}')|$ and partition $\partial_{i^*}(\mathcal{G}')$ into $(\mathcal{F}'_a : a \in [m])$ so that $\sum_a \mu(\mathcal{F}'_a) = \mu(\partial_{i^*}\mathcal{G}') \geqslant \mu(\mathcal{G}')m/hn \geqslant \eta/hn$. By definition of $\mathcal{J}$ and repeated merging, we can form $\mathcal{F}_1, \dots, \mathcal{F}_h$ of the form $\mathcal{F}_i = \cup_{a \in S_i} \mathcal{F}'_a$ with each $\mu(\mathcal{F}_i) \in (\eta^2/n^2, 2\eta^2/n^2)$. However, these cross contain $\mathcal{H}$ by Lemma 7.3. $\square$

9 | CONCLUDING REMARKS

An open problem is to decide whether our main theorem holds for the binary alphabet $m = 2$. Here our junta approximation method cannot work, as the (conjectural) extremal examples are not juntas: they are balls depending on all coordinates. Despite this, it is still plausible that a result can be obtained by a stability method, by adapting the methods of [18] in proving stability for Katona's intersection theorem.

Another natural open problem is to obtain an infinitary version of our main theorem. Say $A \subset \mathbb{R}^n$ is $(t-1)$-avoiding if it contains no pair x, y with $|\{i : x_i = y_i\}| = t - 1$. What is the maximum possible Hausdorff dimension of A? At first, one might think that the answer is $n - t$, and that this would follow from our theorems for large finite alphabets via a standard limiting argument if one assumes that A is closed. One must make some assumption on A for any non-trivial result, as there are pathological examples of $A \subset \mathbb{R}^n$ of Hausdorff dimension n in which any distinct x, y have $x_i \neq y_i$ for all $i \in [n]$. However, even when A is closed, there are some surprises. For example, although it is not hard to see that a 1-avoiding set in $[m]^3$ has size $O(m)$, there is a closed 1-avoiding set $A \subset \mathbb{R}^3$ with Hausdorff dimension 2: this can be achieved by $A = \{(x, f(x), f(x)) : x \in [0, 1]\}$ for a suitably pathological continuous function f.

ACKNOWLEDGEMENTS

We thank Ben Green for helpful remarks regarding the infinitary forbidden intersection problem. P. K. is supported by ERC Advanced Grant 883810.

JOURNAL INFORMATION

The *Journal of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

ORCID

Dor Minzer https://orcid.org/0000-0002-8093-1328

REFERENCES

1. R. Ahlswede and L. H. Khachatrian, *The complete intersection theorem for systems of finite sets*, European J. Combin. **18** (1997), no. 2, 125–136.
2. R. Ahlswede and L. H. Khachatrian, *The diametric theorem in Hamming spaces - optimal anticodes*, Adv. Appl. Math. **20** (1998), no. 4, 429–449.
3. F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer, *Some intersection theorems for ordered sets and graphs*, J. Combin. Theory Ser. A **43** (1986), no. 1, 23–37.
4. I. Dinur and E. Friedgut, *Intersecting families are essentially contained in juntas*, Combin. Probab. Comput. **18** (2009), no. 1-2, 107–122.
5. S. Eberhard, J. Kahn, B. Narayanan, and S. Spirkl, *On symmetric intersecting families of vectors*, arXiv:1909.11578, 2019.
6. D. Ellis, N. Keller, and N. Lifshitz, *Stability for the complete intersection theorem, and the forbidden intersection problem of Erdős and Sós*, arXiv:1604.06135, 2016.
7. D. Ellis, N. Keller, and N. Lifshitz, *Stability versions of Erdős-Ko-Rado type theorems, via isoperimetry*, J. Eur. Math. Soc. **21** (2019), no. 12, 3857–3902.
8. P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, Q. J. Math. **12** (1961), no. 1, 313–320.
9. P. Frankl and Z. Füredi, *The Erdős-Ko-Rado theorem for integer sequences*, SIAM J. Algebraic Discrete Methods **1** (1980), 376–381.
10. P. Frankl and Z. Füredi, *Forbidding just one intersection*, J. Combin. Theory Ser. A **39** (1985), no. 2, 160–176.
11. P. Frankl and V. Rödl, *Forbidden intersections*, Trans. Amer. Math. Soc. **300** (1987), no. 1, 259–286.
12. P. Frankl and N. Tokushige, *The Erdős–Ko–Rado theorem for integer sequences*, Combinatorica **19** (1999), no. 1, 55–63.
13. P. Frankl and N. Tokushige, *Invitation to intersection problems for finite sets*, J. Combin. Theory Ser. A **144** (2016), 157–211.
14. H. Hatami, *A structure theorem for boolean functions with small total influences*, Ann. Math. **176** (2012), 509–533.
15. J. Hązła, T. Holenstein, and E. Mossel, *Product space models of correlation: between noise stability and additive combinatorics*, Discrete Anal. **20** (2018), 63 p.
16. P. Keevash, N. Lifshitz, E. Long, and D. Minzer, *Global hypercontractivity and its applications*, arXiv:2103.04604, 2021.
17. P. Keevash and E. Long, *Frankl-Rodl type theorems for codes and permutations*, Trans. Amer. Math. Soc. **369** (2017), 1147–1162.
18. P. Keevash and E. Long, *Stability for vertex isoperimetry in the cube*, arXiv:1807.09618, 2018.
19. N. Keller and N. Lifshitz, *The junta method for hypergraphs and the Erdős-Chvátal simplex conjecture*, arXiv:1707.02643, 2017.
20. A. Kostochka, D. Mubayi, and J. Verstraëte, *Turán problems and shadows I: paths and cycles*, J. Combin. Theory Ser. A **129** (2015), 57–79.
21. D. A. Levin and Y. Peres, *Markov chains and mixing times*, 2nd ed., American Mathematical Society, Providence, RI, 2017.
22. N. Lifshitz and D. Minzer, *Noise sensitivity on the p-biased hypercube*, 60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9–12, 2019, pp. 1205–1226.

23. E. Mossel, *Gaussian bounds for noise correlation of functions*, Geom. Funct. Anal. **19** (2010), no. 6, 1713–1756.
24. E. Mossel, *Gaussian bounds for noise correlation of resilient functions*, arXiv:1704.04745, 2017.
25. E. Mossel, R. O'Donnell, and K. Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality*, Ann. Math. (2010), 295–341.
26. D. Mubayi and J. Verstraëte, *A survey of Turán problems for expansions*, Recent trends in combinatorics, Springer, Switzerland, 2016, pp. 117–143.
27. R. O'Donnell, *Analysis of boolean functions*, Cambridge University Press, New York, NY, 2014.
28. J. Pach and G. Tardos, *Cross-intersecting families of vectors*, Graphs Combin. **31** (2015), 477–495.