

Moonflowers and code sparsification

Shachar Lovett*
UC San Diego

Raghu Meka†
UCLA

Yimeng Wang‡
UCLA

July 20, 2026

Abstract

We introduce *moonflowers*, a weaker analogue of sunflowers. A family of sets $S_1, \dots, S_k$ is a k -moonflower if each set S_i contains at least one element that is absent from all the others. We study the extremal problem of determining the largest possible size of a family of sets of size at most w that avoids a k -moonflower, and obtain optimal bounds.

Using the extremal moonflower bound and Gilmer’s entropy lemma, we obtain universe reduction theorem for moonflower-free families. As an application, we revisit the code sparsification problem studied by Brakensiek and Guruswami (STOC 2025). Using our extremal bounds for moonflowers, we get near-optimal bounds for code sparsification while also simplifying the argument. Concretely, we improve the dependence on the block length from poly-logarithmic to logarithmic, which is necessary.

1 Introduction

Extremal combinatorics is a branch of combinatorics that studies how large a finite combinatorial object needs to be in order to guarantee the existence of certain patterns of interest. One such pattern that has attracted the attention of researchers in the past few decades is *sunflowers*. A collection of distinct sets $S_1, \dots, S_k$ is called a k -*sunflower* if their pairwise intersections are all the same; in other words, $S_i \cap S_j = S_1 \cap \dots \cap S_k$ for all $i \neq j$.

Setting terminology, a set S is called a w -set if $|S| \leq w$. Back in 1960, Erdős and Rado [ER60] proved that any large family of w -sets must contain a k -sunflower. Specifically, they showed that if $\mathcal{F}$ is a family of w -sets of size $|\mathcal{F}| \geq w! \cdot (k-1)^w$, then $\mathcal{F}$ must contain a k -sunflower. In the same work, Erdős and Rado [ER60] conjectured the bound can be significantly improved.

Conjecture 1.1 (Sunflower conjecture [ER60]). *Let $k \geq 3$. There exists $c = c(k)$ such that any family of w -sets $\mathcal{F}$ of size $|\mathcal{F}| \geq c^w$ must contain a k -sunflower.*

Despite the simplicity of the statement, improving upon the bound in [ER60] turned out to be rather challenging. For almost sixty years after the sunflower conjecture was raised, even for the case of $k = 3$, the best known bound was still of the form $|\mathcal{F}| \geq w^{O(w)}$ [Kos97, Fuk18]. This long period void of significant progress finally ended with the work [ALWZ21] which improved the upper bound to $|\mathcal{F}| \geq (\log w)^w (k \log \log w)^{O(w)}$. Subsequent works [Rao20, BCW21] built upon the result in [ALWZ21] and obtained the following improved bound.

Lemma 1.2 (Improved bounds for the sunflower lemma [BCW21]). *There exists a constant $C > 0$ such that the following holds. Any family of w -sets $\mathcal{F}$ of size $|\mathcal{F}| \geq (Ck \log w)^w$ must contain a k -sunflower.*

*Supported by Simons Investigator Award #929894 and NSF award CCF-2425349

†Supported by NSF EnCORE: Institute for Emerging CORE Methods in Data Science Award #2217033 and NSF AF: Small Award #2425350

‡Supported by NSF EnCORE: Institute for Emerging CORE Methods in Data Science Award #2217033 and NSF AF: Small Award #2425350

One of the key innovations of [ALWZ21] was the use of a robust probabilistic generalization of sunflowers called *robust sunflowers*. They showed that if a set family $\mathcal{F}$ is a robust sunflower of the appropriate parameters, then $\mathcal{F}$ must contain a large sunflower. Using this, they reduced the problem of finding sunflowers to the problem of finding robust sunflowers. Through a careful counting argument, they were able to establish the robust sunflower lemma which leads to the final improved bound for the sunflower lemma.

It is worth noting that the robust sunflower lemma parameters are known to be tight. Hence, any improvement to the sunflower lemma must come from a different approach. Apart from its application in the improved sunflower lemma, robust sunflowers found various applications in combinatorics and theoretical computer science. In combinatorics, it led to the resolution of the famous Kahn-Kalai conjecture [KK07] by Park and Pham [PP24]. In theoretical computer science, it was used in proving monotone circuit lower bounds [Ros14, CKR22] and lifting theorems in communication complexity [LMM⁺22].

In short, robust sunflowers are a generalization of sunflowers for which one can prove tight bounds, and which has diverse applications in combinatorics and theoretical computer science. This motivates the following broad question:

Are there other natural generalizations / variants of sunflowers for which we are able to prove tight bounds and have applications in combinatorics and theoretical computer science?

1.1 Moonflowers

In this paper, we study a combinatorial object which we call *moonflowers*. A family of sets $S_1, \dots, S_k$ is called a *k-moonflower* if each S_i contains at least one element absent from all other sets. Equivalently, there is a set I such that the sets $S_1 \setminus I, \dots, S_k \setminus I$ are nonempty and pairwise disjoint. Note that unlike for sunflowers, we do not require anything on the intersections $S_i \cap I$.

It is easy to see that a sunflower is either a moonflower or becomes one after removing one set¹. The converse, however, is not true. For example, take any set system $\mathcal{F}$ and add a new element to each set; this forms a moonflower but in general not a sunflower.

Moreover, for a fixed set family $\mathcal{F}$, the gap between the size of the largest sunflower it contains and the size of the largest moonflower it contains can be exponentially large. As an example, consider the set family $\mathcal{F} := \{S \subseteq [2n] : |S \cap \{2i-1, 2i\}| = 1, \forall i \in [n]\}$. Then $|\mathcal{F}| = 2^n$ and $\mathcal{F}$ does not contain a 3-sunflower². Next, enlarge the universe to size $2^n + 2n$ and add one new distinct element to each set in $\mathcal{F}$. Denote the resulting family by $\mathcal{F}'$. Then $\mathcal{F}'$ still does not contain a 3-sunflower. However, $\mathcal{F}'$ is a 2^n -moonflower.

Just as in the case of sunflowers, we are interested in the following extremal question on moonflowers:

Question 1.3. *How large does a family of w -sets have to be to guarantee the existence of a k -moonflower?*

We prove the following theorem.

Theorem 1.4 (Extremal bounds on moonflowers). *Let $\mathcal{F}$ be a family of w -sets which is k -moonflower-free. Then*

$$|\mathcal{F}| \leq \binom{k+w-1}{w}.$$

In particular, when $w = \Theta(k)$, the above bound simplifies to $|\mathcal{F}| \leq \exp(O(w))$. We complement this upper bound with a matching lower bound via an explicit construction. First, note that the family of all w -sets in $[k+w-2]$ is k -moonflower-free and has size $\binom{k+w-2}{w}$ which is near optimal. A modification of this construction, found by ChatGPT, gives a tight bound.

Lemma 1.5 (Moonflower lower bound). *There exists a family of w -sets $\mathcal{F}$ such that $|\mathcal{F}| = \binom{k+w-1}{w}$ and $\mathcal{F}$ is k -moonflower-free.*

¹More precisely, let $S_1, \dots, S_k$ be a sunflower and set $K = \cap S_i$. If K is not one of the sets S_i then $S_1, \dots, S_k$ is a moonflower. If it is, then after removing it the remaining $k-1$ sets form a moonflower.

²To see this, suppose $S_1, S_2, S_3 \in \mathcal{F}$ are distinct sets that form a 3-sunflower. For any $i \in [n]$, at least two of the sets must intersect the same element in $\{2i-1, 2i\}$ and hence all three must contain it. This implies that $S_1 = S_2 = S_3$ and so the sets are not distinct.

Consequently, the upper bound in Theorem 1.4 is tight and $\text{MF}(k, w) = \binom{k+w-1}{w}$. Similar to robust sunflowers, the notion of moonflowers has applications in both combinatorics and theoretical computer science. In combinatorics, it is identical to *induced matchings* in a bipartite graph. Specifically, given a graph $G = (V, E)$, a set of edges $M \subseteq E$ is called an *induced matching* of G if (i) E is a matching and (ii) there is no edge in G connecting the endpoints of two distinct edges in M .

Given a set family $\mathcal{F}$, we can treat it as a bipartite graph $G_{\mathcal{F}} = (L_{\mathcal{F}}, R_{\mathcal{F}}, E)$ where vertices in $L_{\mathcal{F}}$ corresponds to sets in $\mathcal{F}$ and vertices in $R_{\mathcal{F}}$ corresponds to $\text{supp}(\mathcal{F})$. Then, it is not difficult to see that a k -moonflower in $\mathcal{F}$ corresponds exactly to an induced matching of size k in $G_{\mathcal{F}}$. Because of this equivalence, Theorem 1.4 immediately implies the following graph-theoretic result.

Corollary 1.6 (Extremal bounds on induced matchings). *Let $G = (L, R, E)$ be a left degree at most w bipartite graph with no isolated vertices such that no two vertices in L have the same neighborhood. If G has no induced matchings of size k , then*

$$|L| \leq \binom{k+w-1}{w}.$$

1.2 Universe Reduction

Let $\mathcal{F} \subseteq 2^{[n]}$ be a set family over universe $[n]$. In general, the universe can be quite large. For the purpose of efficient computations and storage memory, we would like to identify a *small* set of important elements $I \subseteq [n]$ such that we do not lose too many sets in $\mathcal{F}$ after reducing the universe to I , i.e., the size $\mathcal{F}_I := \{S \in \mathcal{F} : S \subseteq I\}$ is not too small compared to that of $\mathcal{F}$. In fact, as we will see in the code sparsification application, what we really need is that after puncturing a few coordinates in I , the residual set system is small. That is, $\mathcal{F}_{\bar{I}} = \{S \setminus I : S \in \mathcal{F}\}$ should be small.

In general, this task is not possible. For instance, consider the set $\mathcal{F} = 2^{[n]}$, i.e., the set of all subsets of $[n]$. Then every element in $[n]$ is contained in exactly half of the sets in $\mathcal{F}$. Leaving out even one element would lose $|\mathcal{F}|/2$ of the sets, and we need to puncture most elements to reduce the size of $\mathcal{F}$ significantly, say to $2^{n/2}$.

Therefore, it would be ideal if we can identify structures over set families that make such universe reduction possible. It turns out moonflower-freeness is one such desirable structure. We first show the existence of a popular element in large moonflower-free set families.

Theorem 1.7 (Large moonflower-free set family implies existence of popular element). *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w -sets which is k -moonflower-free. If $|\mathcal{F}| > \binom{k+tw}{tw}^{\Omega((\log t)/t)}$ for some integer $t \geq 2$, then there exists an element $i \in [n]$ such that i is contained in more than $1/2t$ fraction of the sets in $\mathcal{F}$.*

The proof of the above relies on our extremal bound for moonflowers and Gilmer's entropy lemma [Gil22]. We further show that if a w -set family $\mathcal{F}$ is k -moonflower-free, then we can identify a small set of elements that cover most sets, in the sense that puncturing them results in a small residual set system.

Theorem 1.8 (Universe reduction). *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w -sets which is k -moonflower-free. Then for any $\lambda \in (0, 1]$, there exists $I \subseteq [n]$ of size $|I| \leq O(\lambda^{-1} k \log^3(kw/\lambda))$ satisfying $|\mathcal{F}_{\bar{I}}| \leq \exp(\lambda w)$.*

Because of its simplicity and generality, we believe Theorem 1.8 can be an interesting standalone result with broader applications. We complement this result with a lower bound showing it is near optimal in the interesting regime of parameters $k \gg \lambda w \gg 1$ (we justify why this is the interesting regime in Section 4.5).

Lemma 1.9 (Universe reduction lower bound). *Let $k, w \geq 2$ be positive integers and $\lambda \in (0, 1]$. Assume that $k = \Omega(\lambda w)$ and $\lambda w = \Omega(1)$. Then there exists a family of w -sets $\mathcal{F} \subseteq 2^{[n]}$ which is k -moonflower-free such that, any subset $I \subseteq [n]$ satisfying $|\mathcal{F}_{\bar{I}}| \leq \exp(\lambda w)$ must have size $|I| \geq \Omega(\lambda^{-1} k)$.*

1.3 Code sparsification

One application of our extremal moonflower bounds and universe reduction theorem is *code sparsification*. Code sparsification studies the following question: given an arbitrary code $\mathcal{C} \subseteq \{0, 1\}^n$, can we restrict $\mathcal{C}$ to only a few weighted coordinates while approximately preserving the weight of all codewords?

Formally speaking, let $\mathcal{C} \subseteq \{0, 1\}^n$ be an arbitrary code. A *weighted coordinate set* is a pair (T, α) where $T \subseteq [n]$ and $\alpha : T \rightarrow \mathbb{R}_{\geq 0}$. It induces the estimator

$$\widehat{\text{wt}}_{T, \alpha}(x) := \sum_{i \in T} \alpha(i) x_i, \quad x \in \{0, 1\}^n.$$

We say (T, α) ε -*sparsifies* $\mathcal{C}$ if $\widehat{\text{wt}}_{T, \alpha}(x) \in (1 \pm \varepsilon) \text{wt}(x)$ for all $x \in \mathcal{C}$ where $\text{wt}(x) = \sum_{i=1}^n x_i$. Such a pair (T, α) is called an ε -*sparsifier* of $\mathcal{C}$.

In the case where $\mathcal{C}$ is a linear code, [KPS24, KPS25] showed that in randomized $\text{poly}(n, 1/\varepsilon)$ time, one can compute an ε -sparsifier with support size $|T| \leq \tilde{O}(\dim(\mathcal{C})/\varepsilon^2)$. For general (not necessarily linear) codes, [BG25] extended the results in [KPS24, KPS25] and showed that any code $\mathcal{C} \subseteq \{0, 1\}^n$ admits an ε -sparsifier with support size $|T| \leq O(\text{NRD}(\mathcal{C})(\log n)^6/\varepsilon^2)$ where $\text{NRD}(\mathcal{C})$ denotes the *non-redundancy* of $\mathcal{C}$ defined as follows.

Definition 1.10 (Non-redundancy [BG25]). *A subset $I \subseteq [n]$ is non-redundant for a code $\mathcal{C} \subseteq \{0, 1\}^n$ if for each $i \in I$, there exists $c \in \mathcal{C}$ such that $c_i = 1$ and $c_j = 0$ for all $j \in I \setminus \{i\}$. We define the non-redundancy of $\mathcal{C}$, denoted by $\text{NRD}(\mathcal{C})$, to be the size of the largest set that is non-redundant for $\mathcal{C}$.*

If we view a code $\mathcal{C} \subseteq \{0, 1\}^n$ as a $|\mathcal{C}| \times n$ matrix, then it is clear from the definition that $\text{NRD}(\mathcal{C})$ is the dimension of the largest permutation submatrix contained in $\mathcal{C}$. When $\mathcal{C}$ is a linear code, $\text{NRD}(\mathcal{C})$ equals the dimension of $\mathcal{C}$. More importantly for us, NRD is closely connected to moonflowers.

Given a set family $\mathcal{F}$, let $\text{MF}(\mathcal{F})$ denote the size of the largest moonflower contained in $\mathcal{F}$. Denote $\mathcal{F}_{\mathcal{C}} := \{\text{supp}(c) : c \in \mathcal{C}\} \subseteq 2^{[n]}$. Then we have $\text{MF}(\mathcal{F}_{\mathcal{C}}) = \text{NRD}(\mathcal{C})$. We introduce the new nomenclature instead of using NRD to draw a parallel with the bounds on classical sunflower lemma bounds, which inspired the extremal questions, e.g., Question 1.3, we study.

For a proof of the equivalence, see Lemma 2.8. In particular, this allows us to apply Theorem 1.4 to bound the size of $\mathcal{C}_{\leq d} := \{c \in \mathcal{C} : \text{wt}(c) \leq d\}$ for any $d \in [n]$ as moonflower-freeness is preserved under projections (see Lemma 2.4). Using this and a refined analysis of the techniques introduced in [BG25], we are able to obtain the following result on code sparsification.

Theorem 1.11 (Improved code sparsification). *Let $\mathcal{C} \subseteq \{0, 1\}^n$ with $\text{NRD}(\mathcal{C}) = k$. Then for every $\varepsilon \in (0, 1/4)$ there exists a weighted coordinate set (T, α) that ε -sparsifies $\mathcal{C}$ and satisfies*

$$|T| \leq \frac{k \log n}{\varepsilon^2} \cdot \text{poly}(\log(k/\varepsilon), \log \log n).$$

Comparing with the result in [BG25], we bring the dependence on n to be optimal up to $\text{poly}(\log \log n)$ factors instead of being off by $\text{poly}(\log n)$ factors. In addition, while we build on the work of [BG25], our arguments are simpler owing to the tight moonflower bound (which is also obtained by elementary means).

As the following lemma shows, the $\log n$ dependence is necessary for ε -sparsifiers in general. A modification of an argument from [BG25] shows the following:

Lemma 1.12 (Lower bound on code sparsification). *Let $k \geq 1$ and $\varepsilon \in (0, 1)$. Then, for all large enough n , there exists an explicit $\mathcal{C} \subseteq \{0, 1\}^n$ with $\text{NRD}(\mathcal{C}) = k$ such that any ε -sparsifier (T, α) of $\mathcal{C}$ must satisfy*

$$|T| = \Omega\left(\frac{k \log(n/k)}{\varepsilon}\right).$$

For $k = O(1)$, the above lemma implies $|T| = \Omega((\log n)/\varepsilon)$, matching the upper bound in Theorem 1.11 up to polynomial factors in $1/\varepsilon$. We prove this simple claim in Section 5.2. It is an interesting question to see if the dependence on ε in the lower bound can be improved, or if the upper bound dependency on ε can be improved.

Chain-Length Complexity. A quantity related to NRD called *chain-length* complexity has also been studied in the context of sparsification in [BGP26]. For a set $\mathcal{C} \subseteq \{0, 1\}^n$, let $\text{CL}(\mathcal{C})$ denote the length of the longest ascending chain in the union closure of the set-family $\mathcal{F}_{\mathcal{C}} = \{\text{supp}(c) : c \in \mathcal{C}\}$. Since $\text{NRD}(\mathcal{C}) \leq \text{CL}(\mathcal{C})$, the results of [BG25] also imply sparsification in terms of $\text{CL}(\mathcal{C})$. In [BGP26], the

authors proved a sparsification upper bound of the form $O(\text{CL}(\mathcal{C}) \text{poly log}(\text{CL}(\mathcal{C}))/\varepsilon^2)$. Note that there is no dependence on the universe-size n which stands in contrast to our lower bound above. [BGP26] also show a lower bound of $\Omega(\text{CL}(\mathcal{C})/\varepsilon^2)$ for sparsifying some set systems via connections to data structure lower bounds for graph cuts in [CKST19].

However, NRD can be much smaller than CL in general. If we take a matrix view of a code $\mathcal{C}$, then $\text{CL}(\mathcal{C})$ is the size of the largest upper triangular matrix while $\text{NRD}(\mathcal{C})$ is the size of the largest identity matrix. For instance, take $\mathcal{C} = \{(1, 0, \dots, 0), (1, 1, 0, \dots, 0), \dots, (1, 1, 1, \dots, 1)\} \subseteq \{0, 1\}^n$. Then $\text{NRD}(\mathcal{C}) = 1$, whereas $\text{CL}(\mathcal{C}) = n$. Here, our sparsification upper bound is nearly tight, whereas the bounds based on chain-length are notably worse. This makes our bounds incomparable to those of [BGP26] in general.

1.4 Proof overview

In this subsection, we present the overview of our proofs.

Moonflowers. Frankl [Fra82] showed that if $\mathcal{A} = \{A_1, \dots, A_m\}$ is a family of r -sets and $\mathcal{B} = \{B_1, \dots, B_m\}$ is a family of s -sets such that (1) $A_i \cap B_i = \emptyset$ for $i = 1, 2, \dots, m$ and (2) $A_i \cap B_j \neq \emptyset$ for $1 \leq i < j \leq m$. Then $m \leq \binom{r+s}{s}$.

Given this result, we show that starting from a family of w -sets which is k -moonflower free, we can recursively define set families $\mathcal{A}$ and $\mathcal{B}$ using sets of minimal size and minimal intersections such that the resulting $\mathcal{A}$ and $\mathcal{B}$ satisfy the conditions of [Fra82]. Theorem 1.4 then follows immediately. The resulting proof is a simple reduction to Frankl's classical statement. For more details, see Section 3.

Universe reduction. We first illustrate how we prove the existence of a popular element in a large family of w -sets which is k -moonflower free. To this end, we crucially apply the entropy lemma [Gil22, Saw23] used to obtain improved bounds on the union-closed set-family conjecture [Fra95].

More concretely, let $\mathcal{D}$ be a probability distribution over $2^{[n]}$. Gilmer's entropy lemma states the following: *if there is no popular element under $\mathcal{D}$, then taking unions of i.i.d copies of $A \sim \mathcal{D}$ increases the entropy.* Take $\mathcal{D}$ to be the uniform distribution over $\mathcal{F}$. Using this, we obtain lower bounds on the entropy of a distribution over unions of $\mathcal{U}_t(\mathcal{F})$, the set family consisting of unions of t sets in $\mathcal{F}$. Notice this implies a lower bound on the size of $\mathcal{U}_t(\mathcal{F})$ as entropy is upper bounded by the logarithm of the domain size. On the other hand, since moonflower-freeness is preserved under unions (Lemma 2.5), we can upper bound $|\mathcal{U}_t(\mathcal{F})|$ via the extremal moonflower bound (Theorem 1.4). Combining these two bounds yields Theorem 1.7.

Now we take a step further. Instead of the existence of a single popular element, we want to show the existence of a small set of elements that support most of the sets in $\mathcal{F}$. To this end, consider the following definition. We say a set family $\mathcal{F} \subseteq 2^{[n]}$ is p -covered if there exists a distribution Q on $[n]$ such that for every $S \in \mathcal{F}$, $\Pr_{i \sim Q}[i \in S] \geq p$. Thus, if our family of interest $\mathcal{F}$ is p -covered, then we can simply sample coordinates from Q . As long as we sample enough coordinates, then p -coveredness guarantees most sets in $\mathcal{F}$ are contained in the sampled coordinates.

While the given set family $\mathcal{F}$ may not necessarily be p -covered, by appealing to LP-duality, we show that if we allow removal of a few sets from $\mathcal{F}$, something stronger holds true: for every $J \subseteq [n]$, if we define $\mathcal{F}_J := \{S \cap J : S \in \mathcal{F}\}$, then there exists an exceptional set $\mathcal{S}_J \subseteq \mathcal{F}_J$ of size $|\mathcal{S}_J| \leq M$ such that $\mathcal{F}_J \setminus \mathcal{S}_J$ is p -covered. We say $\mathcal{F}$ is (p, M) -almost-covered if it satisfies this condition.

The last piece of the puzzle is how we can bound the parameter M . Indeed, the validity of the entire argument rests on M being reasonably small. To show this, we crucially rely on the extremal moonflower bound (Theorem 1.4). Combining this with Gilmer's entropy lemma, we are able to obtain good control on M .

We note that the LP-duality and sampling argument were used in [BG25] and a similar universe reduction result was proved. The key component was also Gilmer's entropy lemma. But to establish their result, they appealed to VC-dimension [Sau72, She72]. Our proof is noticeably simpler, and we obtain nearly tight control of the size of the popular set owing in part to the optimal extremal bound.

Code sparsification. We now pivot our discussion to code sparsification. In what follows, we will start with the simplest sparsifier and gradually refine it to obtain our final optimal sparsifier. Given a code $\mathcal{C} \subseteq \{0, 1\}^n$, the simplest sparsifier we can have is the following: *include each coordinate independently with*

probability $1/2$ to form a set T and assign each coordinate a weight of 2. This weighting scheme makes sure the weights are preserved in expectation.

However, this sparsifier has an immediate problem: the weights of the small-weight codewords are not necessarily preserved. For instance, suppose we have a codeword $c \in \mathcal{C}$ with weight $\text{wt}(c) = 1$. Then the only way for c 's weight to be preserved is to include the sole coordinate in $\text{supp}(c)$ to the sparsifier and this happens only with probability $1/2$. This probability quickly goes down as we have more such low-weight codewords in $\mathcal{C}$, deeming this simple sparsifier implausible.

To rectify this, we include the support of all low-weight codewords to the sparsifier. Specifically, set a threshold $w_{\min}$. Let $\mathcal{C}_{\leq w_{\min}}$ denote the set of all codewords in $\mathcal{C}$ with weight $\leq w_{\min}$. Include $\bigcup_{c \in \mathcal{C}_{\leq w_{\min}}} \text{supp}(c)$ into the sparsifier T and assign each such coordinate a weight of 1. Then we add each coordinate $i \in [n] \setminus (\bigcup_{c \in \mathcal{C}_{\leq w_{\min}}} \text{supp}(c))$ to T with probability $1/2$ independently and assign each such coordinate a weight of 2. Every codeword $c \in \mathcal{C}_{w_{\min}}$ has its weight preserved exactly. For $c \in \mathcal{C}$ such that $|c| \geq w_{\min}$, using a standard Chernoff bound, the probability that a codeword c 's weight is not preserved up to an additive ε factor is at most $\exp(-O(\varepsilon^2|c|)) \leq \exp(-O(\varepsilon^2 w_{\min}))$. Taking the union bound over $|\mathcal{C}|$ elements, the probability of some codeword's weight not being preserved is at most $|\mathcal{C}| \exp(-O(\varepsilon^2 w_{\min}))$. Setting $w_{\min} = O(\log |\mathcal{C}| / \varepsilon^2)$ yields a failure probability of $O(1)$.

Despite its viability, the sparsifier above suffers from two major inefficiencies: (i) Since we only set a single threshold $w_{\min}$, we have a single failure probability upper bound for all codewords with weight $\geq w_{\min}$. This can be extremely lossy for codewords with large weights. (ii) Because this simple approach does not have any control on the size $|\mathcal{C}_{\geq w_{\min}}|$, the only thing we can do is to upper bound this quantity using $|\mathcal{C}|$.

In [BG25], the authors propose the following improved sparsifier that overcomes some of the inefficiencies: for the low-weight regime ($|c| \leq w_{\min}$), as before, the support of all low-weight codewords is added to the sparsifier. Now for weights $w \geq w_{\min}$, they split the weights into dyadic weight intervals of the form $[w, 2w]$. They then applied a version of Theorem 1.8 to identify an important set of coordinates for codewords over each such dyadic interval and add the identified coordinates to the sparsifier. Then finally, include each coordinate not yet in the sparsifier independently with probability $1/2$ and with weight 2. This new approach has two important benefits: first, the dyadic intervals allow for better Chernoff bounds as we have a more careful treatment of the weights. Second, for the high-weight codewords, because of the small trace guarantee of the identified coordinates, the union bound is now taking over a much smaller set.

Now we discuss how we further improve upon the sparsifier in [BG25]. Given a code $\mathcal{C} \subseteq \{0, 1\}^n$, consider the set family $\mathcal{F}_{\mathcal{C}} := \{\text{supp}(c) : c \in \mathcal{C}\}$. We make the observation that $\text{NRD}(\mathcal{C}) = \text{MF}(\mathcal{F}_{\mathcal{C}})$ where $\text{MF}(\mathcal{F})$ denotes the size of the largest moonflower contained in set family $\mathcal{F}$. For a proof of this, see Lemma 2.8. We divide the weights into three regimes: low-weight ($|c| \leq w_{\text{low}}$); medium-weight ($w_{\text{low}} \leq |c| \leq w_{\text{high}}$) and high-weight ($|c| \geq w_{\text{high}}$) regimes. For the low-weight regime, we again include the support of all low-weight codewords into the sparsifier. For the medium-weight regime, we again use the dyadic intervals to identify coordinates to add to the sparsifier. However, instead of doing this all the way up to weight n , we stop at some weight $w \leq w_{\text{high}}$ and include the other coordinates independently with probability $1/2$. The reason for this is the following: the trace bound we obtain from Theorem 1.8 deteriorates as the weight increases. At that point, the bound $|\mathcal{F}_{\leq w}| \leq \binom{k+w}{w}$ we obtain from the optimal moonflower bound in Theorem 1.4 outperforms the bound in Theorem 1.8.

One may argue that we are not including enough coordinates as we could from the dyadic intervals as in [BG25]. However, this would not be an issue as we iterate the entire procedure above on the randomly sampled coordinates S . With high probability, the sampled coordinates S in each iteration has its size halved. Hence, we need to iterate at most $\log n$ iterations and in fact, this is precisely where the $\log n$ factor occurs in our final sparsification result. Finally, we assign weights to coordinates based on the iteration at which they get included in the sparsifier.

Using our universe-reduction bound Theorem 1.8 and the moonflower bound Theorem 1.4 with appropriately chosen parameters, we are able to obtain a sparsifier with the guarantees in Theorem 1.11. For a more thorough discussion of our improved sparsification strategy, see Section 5.1.

Remark. An earlier version of the paper obtained similar results but with more complicated proofs. This version simplifies and streamlines many of the proofs, in particular the extremal bound for moonflowers and universe reduction.

1.5 Organization

In Section 2, we define the relevant terms and prove some elementary results used in the proof of moonflower bound and code sparsification. In Section 3, we prove tight upper and lower bounds for the size of families of w -sets without moonflowers. In Section 4, we prove our main universe reduction theorem. In Section 5, we use this tight extremal bound and universe reduction theorem to prove our code sparsification theorem.

1.6 Future directions

- (1) **From existence to algorithms.** Our results for both moonflowers and code sparsification are existential. A natural follow-up question is: can we make our existence results algorithmic?
- (2) **Better sparsification lower bound.** Our current code sparsification lower bound is near optimal in terms of its dependence on $\text{NRD}(\mathcal{C})$ and n . However, it is not optimal in terms of the dependence on ε . We believe that the correct dependency on the error should be $1/\varepsilon^2$. In addition, it would be nice to have matching lower bound in the interesting special case where the code $\mathcal{C}$ is roughly balanced.
- (3) **More applications of moonflowers.** What other applications are there for moonflowers in combinatorics and theoretical computer science? Given the optimal bounds and the general structure of a moonflower, we believe moonflowers should have a broader range of applications.

AI acknowledgments. While ChatGPT5.5-Pro was used in checking some of the calculations (especially for the somewhat tedious ones in code sparsification), all of the ideas in the paper are due to the authors. The only exception is Lemma 1.5 where ChatGPT improved our lower bound on $\text{MF}(k, w)$ from $\binom{k+w-2}{w}$ to $\binom{k+w-1}{w}$ by modifying our initial construction. ChatGPT5.5-Pro was also used in polishing parts of the writeup but the writing is mostly done by the authors.

Acknowledgments. The authors would like to thank anonymous FOCS reviewers for valuable comments.

2 Preliminaries

In this subsection, we state some elementary definitions and results that will be used throughout this paper. Throughout the paper, $\log(\cdot)$ denotes base-2 logarithms and $\ln(\cdot)$ denotes natural logarithms.

2.1 Sets

Let $[n]$ denote the finite set $[n] := \{1, 2, \dots, n\}$. Throughout this paper, unless otherwise stated, all sets are finite. We use normal letters like S, T to denote sets and calligraphic letters like $\mathcal{F}$ to denote family of sets. For $I \subseteq [n]$, write $\bar{I} := [n] \setminus I$. We use $2^{[n]}$ to denote the family of all subsets of $[n]$.

For $\mathcal{F} \subseteq 2^{[n]}$, define $\text{supp}(\mathcal{F}) := \bigcup_{S \in \mathcal{F}} S$. We say $S \subseteq [n]$ is a w -set if $|S| \leq w$.

Definition 2.1 (Moonflower). *A family of sets $S_1, \dots, S_k \subseteq [n]$ is a k -moonflower if there exists $I \subseteq [n]$ such that the sets $S_i \setminus I$ are all nonempty and pairwise disjoint. We refer to the I of the smallest size satisfying this condition as the core of the moonflower and $S_1, \dots, S_k$ as the petals. A family $\mathcal{F} \subseteq 2^{[n]}$ is k -moonflower-free if it contains no k -moonflower.*

Remark. It is worth noting that the core of a k -moonflower is unique: it is the set of all elements appearing in at least two of the sets S_i .

Let $\mathcal{F}$ be a family of sets. Define $\text{MF}(\mathcal{F})$ to be the largest k such that $\mathcal{F}$ contains a k -moonflower.

Definition 2.2 (Extremal function). *Let $\text{MF}(k, w)$ denote the maximum size of a k -moonflower-free family of w -sets $\mathcal{F}$ (over all universes $[n]$), i.e.,*

$$\text{MF}(k, w) := \max_{\substack{\mathcal{F}: \text{a family of } w\text{-sets} \\ \text{s.t. } \text{MF}(\mathcal{F}) < k}} |\mathcal{F}|$$

Definition 2.3 (Trace). Let $\mathcal{F} \subseteq 2^{[n]}$ be a set family and let $I \subseteq [n]$. The trace of $\mathcal{F}$ on I , denoted $\mathcal{F}_I$ is

$$\mathcal{F}_I := \{S \cap I : S \in \mathcal{F}\}.$$

In words, it is the family obtained by restricting every set $S \in \mathcal{F}$ to the coordinates in I with duplicate sets removed. We will be interested in particular in the trace after puncturing the coordinates in I ,

$$\mathcal{F}_{\bar{I}} := \{S \setminus I : S \in \mathcal{F}\}.$$

One important property of k -moonflower-freeness is that it is preserved under taking projections:

Lemma 2.4 (Moonflower-freeness is preserved under projection). If $\mathcal{F} \subseteq 2^{[n]}$ is k -moonflower-free and $J \subseteq [n]$, then $\mathcal{F}_J$ is k -moonflower-free.

Proof. Suppose for contradiction that $\mathcal{F}_J$ contains a k -moonflower: there exist $S'_1, \dots, S'_k \in \mathcal{F}_J$ and a core $I' \subseteq J$ such that the sets $S'_i \setminus I'$ are nonempty and pairwise disjoint. Choose $S_i \in \mathcal{F}$ with $S_i \cap J = S'_i$. Let $I := I' \cup ([n] \setminus J)$. Then $S_i \setminus I = (S_i \cap J) \setminus I' = S'_i \setminus I'$, hence nonempty and pairwise disjoint. So $S_1, \dots, S_k$ form a k -moonflower in $\mathcal{F}$, contradiction. $\square$

For a set family $\mathcal{F}$, define its t -fold union, denoted $\mathcal{U}_t(\mathcal{F})$ as

$$\mathcal{U}_t(\mathcal{F}) := \{S_1 \cup \dots \cup S_t : S_1, \dots, S_t \in \mathcal{F}\}.$$

It turns out taking t -fold unions also preserves moonflower-freeness.

Lemma 2.5 (Moonflower-freeness is preserved under unions). If $\mathcal{F}$ is k -moonflower-free, then $\mathcal{U}_t(\mathcal{F})$ is k -moonflower-free for all positive integers t .

Proof. Suppose for contradiction that $\mathcal{U}_t(\mathcal{F})$ contains a k -moonflower $A_1, \dots, A_k$ where $A_i = \cup_{\ell=1}^t S_{i,\ell}$ with $S_{i,\ell} \in \mathcal{F}$. For each $i \in [k]$, choose an element $x_i \in A_i$ that is not contained in $\cup_{j \neq i} A_j$ (such an element exists because $A_1, \dots, A_k$ form a moonflower). In particular, x_i must be contained in some S_{i,ℓ_i} . Then $S_{1,\ell_1}, \dots, S_{k,\ell_k}$ form a k -moonflower in $\mathcal{F}$, which is a contradiction. $\square$

Lemma 2.6 (Support bound from moonflower-freeness). Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w -sets which is k -moonflower free. Then

$$|\text{supp}(\mathcal{F})| \leq (k-1)w.$$

In particular, after deleting unused coordinates, we may assume $n \leq (k-1)w$.

Proof. If $\mathcal{F} = \emptyset$ then $\text{supp}(\mathcal{F}) = \emptyset$ and we are done. Let $U := \text{supp}(\mathcal{F})$. Pick a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ that is minimal (under inclusion) subject to $\bigcup_{S \in \mathcal{F}'} S = U$. Then for every $S \in \mathcal{F}'$ there exists an element

$$x_S \in S \setminus \bigcup_{T \in \mathcal{F}', T \neq S} T,$$

otherwise S could be removed while still covering U , contradicting minimality. In particular, the elements $\{x_S : S \in \mathcal{F}'\}$ are pairwise distinct.

Let $I := [n] \setminus \{x_S : S \in \mathcal{F}'\}$. For each $S \in \mathcal{F}'$ we have $S \setminus I = \{x_S\}$, so the sets in $\mathcal{F}'$ form a $|\mathcal{F}'|$ -moonflower with core I . Since $\mathcal{F}$ is k -moonflower-free, we get $|\mathcal{F}'| \leq k-1$. Therefore

$$|U| = \left| \bigcup_{S \in \mathcal{F}'} S \right| \leq \sum_{S \in \mathcal{F}'} |S| \leq |\mathcal{F}'| \cdot w \leq (k-1)w,$$

as required. $\square$

2.2 Codes

We define a Boolean code to be an arbitrary $\mathcal{C} \subseteq \{0, 1\}^n$. We say $\mathcal{C}$ is *non-trivial* if $\mathcal{C} \neq \emptyset, \{0, 1\}^n$. For any $c \in \mathcal{C}$, denote $\text{supp}(c) \subseteq [n]$ to be the set of nonzero coordinates. Define $\text{supp}(\mathcal{C}) := \bigcup_{c \in \mathcal{C}} \text{supp}(c)$. For any codeword $c \in \mathcal{C}$, we define its *Hamming weight* as $|c| := c_1 + \dots + c_n$, i.e., the number of nonzero coordinates. For any $d \in [n]$, let $\mathcal{C}_{\leq d}$ be the set of codewords in $\mathcal{C}$ with Hamming weight at most d .

Given $S \subseteq [n]$ and $c \in \{0, 1\}^n$, we define $c|_S \in \{0, 1\}^S$ to be the list $(c_i : i \in S)$. Likewise, we define *punctured code* $\mathcal{C}|_S := \{c|_S : c \in \mathcal{C}\} \subseteq \{0, 1\}^S$.

Definition 2.7 (Non-redundancy [BG25]). *A subset $I \subseteq [n]$ is non-redundant for a code $\mathcal{C} \subseteq \{0, 1\}^n$ if for each $i \in I$, there exists $c \in \mathcal{C}$ such that $c_i = 1$ and $c_j = 0$ for all $j \in I \setminus \{i\}$. We define the non-redundancy of $\mathcal{C}$, denoted by $\text{NRD}(\mathcal{C})$, to be the size of the largest non-redundant set that is non-redundant for $\mathcal{C}$.*

If we view a code $\mathcal{C}$ as a $|\mathcal{C}| \times n$ matrix, then it is clear from the definition that $\text{NRD}(\mathcal{C})$ is the dimension of the largest permutation submatrix contained in $\mathcal{C}$. When $\mathcal{C}$ is a linear code, $\text{NRD}(\mathcal{C})$ equals the dimension of $\mathcal{C}$. In addition, for any non-trivial $\mathcal{C}$, we have $1 \leq \text{NRD}(\mathcal{C}) \leq n$. Let $\mathcal{F}_{\mathcal{C}} := \{\text{supp}(c) : c \in \mathcal{C}\} \subseteq 2^{[n]}$. We have the following simple lemma establishing a connection between $\text{NRD}(\mathcal{C})$ and $\text{MF}(\mathcal{F}_{\mathcal{C}})$.

Lemma 2.8 (NRD equals MF of the support family). *Let $\mathcal{C} \subseteq \{0, 1\}^n$ be a binary code. Then*

$$\text{MF}(\mathcal{F}_{\mathcal{C}}) = \text{NRD}(\mathcal{C}).$$

Proof. We prove the two inequalities.

(1) $\text{MF}(\mathcal{F}_{\mathcal{C}}) \geq \text{NRD}(\mathcal{C})$. Let $I \subseteq [n]$ be non-redundant for $\mathcal{C}$ with $|I| = \text{NRD}(\mathcal{C})$. By Definition 2.11, for every $i \in I$ there exists a codeword $c^{(i)} \in \mathcal{C}$ such that for all $i' \in I$,

$$c_{i'}^{(i)} = 1 \iff i' = i,$$

equivalently, $\text{supp}(c^{(i)}) \cap I = \{i\}$. Let $S_i := \text{supp}(c^{(i)}) \in \mathcal{F}_{\mathcal{C}}$ and set the core to be $J := [n] \setminus I$. Then for each $i \in I$,

$$S_i \setminus J = S_i \cap I = \{i\},$$

so the sets $\{S_i : i \in I\}$ form a $|I|$ -moonflower (their petals are the singletons $\{i\}$, hence nonempty and pairwise disjoint). Therefore $\text{MF}(\mathcal{F}_{\mathcal{C}}) \geq |I| = \text{NRD}(\mathcal{C})$.

(2) $\text{NRD}(\mathcal{C}) \geq \text{MF}(\mathcal{F}_{\mathcal{C}})$. Let $k := \text{MF}(\mathcal{F}_{\mathcal{C}})$, so $\mathcal{F}_{\mathcal{C}}$ contains a k -moonflower $S_1, \dots, S_k$ with some core $J \subseteq [n]$; write $P_t := S_t \setminus J$ for the petals. By definition, each P_t is nonempty and the P_t 's are pairwise disjoint. Choose an element $i_t \in P_t$ for each $t \in [k]$ and let $I := \{i_1, \dots, i_k\}$. For each t , let $c^{(t)} \in \mathcal{C}$ be a codeword with $\text{supp}(c^{(t)}) = S_t$. We claim that $\text{supp}(c^{(t)}) \cap I = \{i_t\}$. Indeed, $i_t \in S_t$ by construction. If $s \neq t$, then $i_s \notin J$ (since $i_s \in P_s$) and also $i_s \notin S_t \setminus J = P_t$ (since the petals are disjoint), hence $i_s \notin S_t$. Thus $S_t \cap I = \{i_t\}$, meaning that on the coordinate set I , the codeword $c^{(t)}$ has a single 1 exactly at i_t .

Therefore, for every $i_t \in I$ there exists $c^{(t)} \in \mathcal{C}$ whose restriction to I equals the unit vector at i_t . This is exactly the non-redundancy condition of Definition 2.11, so I is non-redundant for $\mathcal{C}$. Hence $\text{NRD}(\mathcal{C}) \geq |I| = k = \text{MF}(\mathcal{F}_{\mathcal{C}})$.

Combining (1) and (2) yields $\text{MF}(\mathcal{F}_{\mathcal{C}}) = \text{NRD}(\mathcal{C})$. $\square$

2.3 Entropy amplification

Definition 2.9 (p -smooth distribution). *A probability distribution D supported on $2^{[n]}$ is p -smooth if for every coordinate $i \in [n]$,*

$$\Pr_{S \sim D}[i \in S] \leq p.$$

Lemma 2.10 (Gilmer's entropy lemma [BG25, Corollary 4.9]). *Let D be a p -smooth distribution over $2^{[n]}$. Let t be a power of two such that $1 - (1 - p)^{t/2} \leq \frac{3 - \sqrt{5}}{2}$ and $A_1, \dots, A_t$ be i.i.d samples from $\mathcal{D}$. Then*

$$H\left(\bigvee_{i=1}^t A_i\right) \geq H(D) \cdot \frac{h(1 - (1 - p)^t)}{h(p)}.$$

Here $h(\cdot)$ is the binary entropy function.

Note that if $p = \Theta(1/t)$, then $h(1 - (1 - p)^t)/h(p) = \Theta(t/\log t)$.

Lemma 2.11 (Entropy bound from size of r -fold unions, variant of [BG25, Corollary 4.13]). *Let $\mathcal{F} \subseteq 2^{[n]}$ be any finite family. For $p \in (0, 1)$, let D be a p -smooth distribution supported on $\mathcal{F}$. If $r \geq 2/p$, then*

$$H(D) \leq O(p \log(1/p) \cdot \log |\mathcal{U}_r(\mathcal{F})|),$$

where $H(\cdot)$ denotes the Shannon entropy.

For a proof of the statement, see [BG25, Corollary 4.13].

2.4 Chernoff bound

We need the following version of Chernoff bound for our sparsification result.

Lemma 2.12 (Chernoff bound, convenient form). *Let $t \geq 1$ and let $X \sim \text{Bin}(t, 1/2)$. Then for every $\Delta > 0$,*

$$\Pr[|2X - t| > \Delta] \leq 2 \exp\left(-\frac{\Delta^2}{3t}\right).$$

Proof. Standard Chernoff bound in additive form. □

3 Extremal moonflower bound

In this section, we prove our extremal moonflower bound.

Theorem 3.1 (Theorem 1.4, restated). *Let $\mathcal{F}$ be a family of w -sets which is k -moonflower-free. Then $|\mathcal{F}| \leq \binom{k+w-1}{w}$.*

Towards the proof of this theorem, we need the following result of Frankl [Fra82].

Theorem 3.2 ([Fra82]). *Let $\mathcal{A} = \{A_1, \dots, A_m\}$ be a family of r -sets and $\mathcal{B} = \{B_1, \dots, B_m\}$ be a family of s -sets such that (1) $A_i \cap B_i = \emptyset$ for $i = 1, 2, \dots, m$ and (2) $A_i \cap B_j \neq \emptyset$ for $1 \leq i < j \leq m$. Then*

$$m \leq \binom{r+s}{s}.$$

We include a proof of Theorem 3.2 in Section A for completeness.

Proof of Theorem 3.1. Let $\mathcal{F}$ be a family of w -sets which is k -moonflower free, and let $m = |\mathcal{F}|$. Let $U := \bigcup_{S \in \mathcal{F}} S$ denote the universe. We construct families $\mathcal{A}$ and $\mathcal{B}$ as follows. Let $\mathcal{F}_1 = \mathcal{F}$, $\mathcal{A} = \mathcal{B} = \emptyset$. For $i = 1, 2, \dots, m$ do:

- Take $B_i \in \mathcal{F}_i$ of minimal size. In particular, B_i does not contain any other set in $\mathcal{F}_i$.
- Let $A_i \subseteq U \setminus B_i$ (not necessarily a set in $\mathcal{F}_i$) be a set of minimal size intersecting all sets in $\mathcal{F}_i \setminus \{B_i\}$. Such a set exists because B_i does not contain any other set in $\mathcal{F}_i$.
- Add A_i to $\mathcal{A}$ and B_i to $\mathcal{B}$.
- Set $\mathcal{F}_{i+1} := \mathcal{F}_i \setminus \{B_i\}$ and repeat.

Denote the resulting set families as $\mathcal{A} = \{A_1, \dots, A_m\}$ and $\mathcal{B} = \{B_1, \dots, B_m\}$. The resulting families $\mathcal{A}$ and $\mathcal{B}$ satisfy the following properties:

- (i) $A_i \cap B_i = \emptyset$ for all $i = 1, 2, \dots, m$.
- (ii) $A_i \cap B_j \neq \emptyset$ for $1 \leq i < j \leq m$.
- (iii) $|A_i| \leq k - 1$ for all $i = 1, 2, \dots, m$.

Notice that (i) holds as A_i is contained in $U \setminus B_i$, (ii) holds because A_i intersects all $S \in \mathcal{F}_i \setminus \{B_i\}$ by definition, which implies that $A_i \cap B_j \neq \emptyset$ for all $j > i$. It remains to show (iii).

Since A_i is minimal, removing any element from A_i would miss some set; that is, for each $s \in A_i$, there exists a set $T_s \in \mathcal{F}_i$ such that $A_i \cap T_s = \{s\}$. This implies that $\{T_s\}_{s \in A_i}$ form a $|A_i|$ -moonflower which is a contradiction if $|A_i| \geq k$.

Consequently, we have obtained a $(k-1)$ -set family $\mathcal{A}$ and a w -set family $\mathcal{B}$ satisfying the conditions of Theorem 3.2. Applying Theorem 3.2, we obtain

$$|\mathcal{F}| \leq \binom{k+w-1}{w}.$$

□

We next prove that the bound is tight.

Lemma 3.3 (Lemma 1.5, restated). *There exists a family of w -sets $\mathcal{F}$ such that $|\mathcal{F}| = \binom{k+w-1}{w}$ and $\mathcal{F}$ is k -moonflower-free.*

Proof. For $s = 0, \dots, w$ define the following family of sets

$$\mathcal{F}_s := \{S \subseteq [k+s-2] : |S| = s\}.$$

Let $\mathcal{F} = \cup_{s=0}^w \mathcal{F}_s$. We claim that $\mathcal{F}$ satisfies the requirement of the lemma. First, note that $\mathcal{F}$ is a family of w -sets of size

$$|\mathcal{F}| = \sum_{s=0}^w \binom{k+s-2}{s} = \binom{k+w-1}{w}.$$

We next show that $\mathcal{F}$ does not contain a k -moonflower. Assume towards a contradiction that it contains a k -moonflower $S_1, \dots, S_k$. For each $i = 1, \dots, k$ let x_i be an element of S_i which is not in any other S_j . Assume without loss of generality that x_1 is maximal among $x_1, \dots, x_k$, and let $s = |S_1|$. Note that $S_1 \subset [k+s-2]$ is a set of size s , and $x_2, \dots, x_k \in [k+s-2]$ are $k-1$ elements distinct from it. This leads to a contradiction. □

4 Universe reduction

In this section, we use the extremal moonflower bound (Theorem 3.1) to obtain two structural results on moonflower-free sets. The main technical ingredient of the proofs is *entropy amplification* [Gil22, Saw23]. As a warm-up, in Section 4.1, using entropy amplification, we show that any large k -moonflower-free set family must contain a popular element. Then, for the rest of this section, we work our way towards proving our main universe reduction theorem (Theorem 1.8).

Before we start with the proof, we need to make a definition that will play an important role in the proof of Theorem 1.8.

Definition 4.1 (p -covered family). *A family $\mathcal{G} \subseteq 2^{[n]}$ is p -covered if there exists a distribution Q on $[n]$ such that for every $S \in \mathcal{G}$,*

$$\Pr_{i \sim Q}[i \in S] \geq p.$$

Equivalently, $\sum_{i \in S} Q(i) \geq p$ for all $S \in \mathcal{G}$.

It is worth noting that the distribution Q in a p -covered family is the LP dual of a p -smooth distribution D . Moreover, if we know a p -cover Q for a set family $\mathcal{F}$ exists, we can simply sample from Q . The main steps of the proof of Theorem 1.8 can then be described as follows:

1. We first show that any p -smooth distribution D supported on a k -moonflower free family $\mathcal{F}$ has entropy at most $O(\log \binom{k+w/p}{k} \cdot p \log(1/p))$. This step critically uses the entropy-amplification property of taking unions proved by [Gil22, Saw23]. See Section 4.4.

2. We then use LP-duality to show the following. Suppose $\mathcal{F}$ has the property that for any $J \subseteq [n]$, any p -smooth distribution on $\mathcal{F}_J$ has entropy at most H . Then, there exists a small set of bad elements $\mathcal{S} \subset \mathcal{F}$, $|\mathcal{S}| \leq 2^H$ such that $\mathcal{F} \setminus \mathcal{S}$ is p -covered. See Section 4.3.
3. Given this, we show that if a family of sets $\mathcal{F}$ has the following property: for any $J \subseteq [n]$, there exists $\mathcal{S}_J \subset \mathcal{F}_J$ of size $|\mathcal{S}_J| \leq 2^H$ such that $\mathcal{F}_J \setminus \mathcal{S}_J$ is p -covered. Then, we can find a small set $I \subseteq [n]$ supporting most of the sets in $\mathcal{F}$. The size of I depends on the quantity H . Theorem 1.8 then follows by bounding the quantity H using extremal moonflower bound (Theorem 3.1). See Section 4.4
4. Finally, in Section 4.5, we complement our universe reduction result with a lower bound.

Remark. The idea of using LP-duality and sampling can be found in [BG25]. With our new extremal moonflower bound (Theorem 3.1), we are able to significantly simplify the proof in [BG25].

4.1 Large family size implies existence of popular elements

Theorem 4.2 (Theorem 1.7, restated). *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w -sets which is k -moonflower-free. If $|\mathcal{F}| > \binom{k+tw}{tw}^{\Omega((\log t)/t)}$ for some integer $t \geq 2$, then there exists an element $i \in [n]$ such that i is contained in more than $1/2t$ fraction of the sets in $\mathcal{F}$.*

The following corollary simplifies the binomial coefficient involved by breaking it down into three cases depending on the relative magnitude of k and tw .

Corollary 4.3. *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w -sets which is k -moonflower-free. Assume for some integer $t \geq 2$ that all elements $i \in [n]$ are contained in less than a $1/2t$ fraction of sets in $\mathcal{F}$. Then*

$$|\mathcal{F}| \leq \begin{cases} \left(\frac{k}{tw}\right)^{O(w \log t)} & \text{when } k \gg tw \\ t^{O(w)} & \text{when } k \asymp tw \\ \left(\frac{tw}{k}\right)^{O(k \log t/t)} & \text{when } k \ll tw. \end{cases}$$

Proof. Use the standard approximation for binomial coefficients. □

Towards a proof of Theorem 4.2, we first use Lemma 2.10 to prove the following corollary that bounds the size of t -fold unions provided the uniform distribution over $\mathcal{F}$ is p -smooth. Recall that $\mathcal{U}_t(\mathcal{F}) := \{S_1 \cup \dots \cup S_t : S_1, \dots, S_t \in \mathcal{F}\}$ is the t -fold union of $\mathcal{F}$.

Lemma 4.4 (Smooth uniform distribution implies large union size). *Let $\mathcal{F}$ be a finite set family. Suppose the uniform distribution over $\mathcal{F}$ is $(1/2t)$ -smooth, then for some constant c_1 , we have*

$$|\mathcal{U}_t(\mathcal{F})| \geq |\mathcal{F}|^{c_1 t / \log t}.$$

Proof. For a fixed t that is a power of 2, let $p = c/t$ for some constant³ $c \leq 0.7$. Let D be the uniform distribution on $\mathcal{F}$, and let $A_1, \dots, A_t \sim D$ be uniformly chosen. Apply Lemma 2.10 to get

$$\log |\mathcal{U}_t(\mathcal{F})| \geq H\left(\bigvee_{i=1}^t A_i\right) \geq H(D) \cdot \frac{h(1 - (1-p)^t)}{h(p)} \geq \log |\mathcal{F}| \cdot \frac{c_1 t}{\log t}$$

for some constant c_1 . Here we used the fact that the distribution $\bigvee_{i=1}^t A_i$ is supported on $\mathcal{U}_t(\mathcal{F})$, and so its entropy is bounded by $\log |\mathcal{U}_t(\mathcal{F})|$, and that D is uniform on $\mathcal{F}$ and so its entropy equals $\log |\mathcal{F}|$. Raising both sides to the power of 2 and simplifying, we obtain the desired bound. □

Lemma 4.5. *Let $\mathcal{F}$ be a family of w -sets which is k -moonflower-free. Suppose the uniform distribution over $\mathcal{F}$ is $(1/2t)$ -smooth, then for some absolute constant $C > 0$ we have*

$$|\mathcal{F}| \leq \binom{k+tw}{tw}^{C(\log t)/t}.$$

³need this to ensure the condition $1 - (1-p)^{t/2} \leq \frac{3-\sqrt{5}}{2}$.

Proof. On one hand, by Lemma 2.5, we know $\mathcal{U}_t(\mathcal{F})$ is a family of tw -sets and k -moonflower free. By Theorem 3.1, we know

$$|\mathcal{U}_t(\mathcal{F})| \leq \binom{k+tw-1}{tw}.$$

On the other hand, by Lemma 4.4, we have $|\mathcal{U}_t(\mathcal{F})| \geq |\mathcal{F}|^{c_1 t / \log t}$. We thus find

$$\binom{k+tw-1}{tw} \geq |\mathcal{F}|^{c_1 t / \log t}.$$

The desired statement then follows by rearranging. $\square$

Proof. (of Theorem 4.2) By Lemma 4.5, we know if $|\mathcal{F}| \geq \binom{k+tw}{tw}^{C(\log t)/t}$ for some large enough constant C , then the uniform distribution over $\mathcal{F}$ cannot be $(1/2t)$ -smooth. This implies there must exist some element $i \in [n]$ that belongs to at least a $(1/2t)$ -fraction of the sets in $\mathcal{F}$. $\square$

4.2 Entropy bound

In this subsection, show the following entropy bound on p -smooth distributions over k -moonflower-free w -set families.

Theorem 4.6 (Entropy bound on p -smooth distributions). *Let $\mathcal{F}$ be a k -moonflower-free family of w -sets. Let D be a p -smooth distribution supported on $\mathcal{F}$ for $0 < p \leq 1/2$. Then*

$$H(D) \leq O\left(p \log(1/p) \cdot \log\left(\binom{k+w\lceil 2/p \rceil}{k}\right)\right).$$

Proof. Let $r = \lceil 2/p \rceil$. Consider the r -fold union family $\mathcal{U}_r(\mathcal{F})$. By Lemma 2.11, we know

$$H(D) \leq O(p \log(1/p) \cdot \log |\mathcal{U}_r(\mathcal{F})|).$$

It remains to bound $|\mathcal{U}_r(\mathcal{F})|$. Notice that every set in $\mathcal{U}_r(\mathcal{F})$ has size at most wr . Moreover, by Lemma 2.5, we know $\mathcal{U}_r(\mathcal{F})$ is k -moonflower-free. We can then apply the extremal moonflower bound (Theorem 3.1) and bound

$$|\mathcal{U}_r(\mathcal{F})| \leq \binom{k+wr}{k}.$$

The desired statement follows. $\square$

4.3 From entropy to coverability

In this section, we show how to obtain a p -cover of $\mathcal{F}$. Namely, we are going to prove the following theorem.

Theorem 4.7 (Coverability from entropy-bounded smoothness, adapted from [BG25, Lemma 4.15]). *Let $\mathcal{F} \subseteq 2^{[n]}$ and $p \in (0, 1)$. Assume every p -smooth distribution supported on $\mathcal{F}$ has entropy $\leq H$. Then there exists $\mathcal{S} \subseteq \mathcal{F}$ with $|\mathcal{S}| \leq 2^H$ such that $\mathcal{F} \setminus \mathcal{S}$ is p -covered.*

While the statement is similar to [BG25, Lemma 4.15], we provide a simplified proof. First, we establish the duality of coveredness and smoothness.

Proposition 4.8 (Cover vs. smooth duality). *Let $\mathcal{G} \subseteq 2^{[n]}$ and $p \in (0, 1)$. Define*

$$\Phi(\mathcal{G}) := \max_{Q \text{ distribution on } [n]} \min_{T \in \mathcal{G}} \sum_{i \in T} Q(i).$$

Then

$$\Phi(\mathcal{G}) = \min_{D \text{ distribution on } \mathcal{G}} \max_{i \in [n]} \mathbf{Pr}[i \in T].$$

In particular,

1. $\mathcal{G}$ is p -covered if and only if $\Phi(\mathcal{G}) \geq p$;
2. If $\mathcal{G}$ is not p -covered, then there exists a p -smooth distribution supported on $\mathcal{G}$.

Proof. Consider the following finite zero-sum game: the row player chooses $T \in \mathcal{G}$ and the column player chooses $i \in [n]$, with payoff matrix $A(T, i) := \mathbf{1}[i \in T]$ to the column player. A mixed strategy for the column player is a distribution Q on $[n]$, and its expected payoff against a pure row choice T is $\sum_{i \in T} Q(i)$. The row player then respond with the worst expected payoff $\min_{T \in \mathcal{G}} \sum_{i \in T} Q(i)$. Maximizing over all mixed strategy Q gives $\Phi(\mathcal{G})$.

A mixed strategy for the row player is a distribution D on $\mathcal{G}$, and the expected payoff against a pure column choice i is $\Pr_{T \sim D}[i \in T]$. The column player best-responds with $\max_i \Pr_{T \sim D}[i \in T]$. The row player then minimizes this quantity.

By von Neumann's minimax theorem [Neu28],

$$\max_Q \min_{T \in \mathcal{G}} \sum_{i \in T} Q(i) = \min_D \max_{i \in [n]} \Pr_{T \sim D}[i \in T],$$

as desired. $\square$

Recall our goal is to obtain a p -cover of $\mathcal{F}$. If our $\mathcal{F}$ is already p -covered, then we are done. Otherwise, by Proposition 4.8 there has to exists a p -smooth distribution on $\mathcal{F}$. Given such a p -smooth distribution, our next Lemma shows how we can obtain a p -covered subfamily from $\mathcal{F}$ by removing only a few sets provided every p -smooth distribution on $\mathcal{F}$ does not have too much entropy.

Proof of Theorem 4.7. If $\mathcal{F}$ is already p -covered, take $S = \emptyset$ and we are done. Otherwise by Proposition 4.8 (2), there exists at least one p -smooth distribution over $\mathcal{F}$. Let

$$\mathcal{P} := \{\nu \text{ distribution on } \mathcal{F} : \nu \text{ is } p\text{-smooth}\}.$$

Then $\mathcal{P}$ is nonempty, compact and convex. Define $\|\nu\|_\infty := \max_{T \in \mathcal{F}} \nu(T)$. This is valid since $\mathcal{F}$ is finite. Let $\tau^* := \min_{\nu \in \mathcal{P}} \|\nu\|_\infty$. By compactness, the minimum is attained. Fix an optimizer $\nu^* \in \mathcal{P}$ and define

$$\mathcal{S} := \{T \in \mathcal{F} : \nu^*(T) = \tau^*\}.$$

We claim that $\mathcal{S}$ is a family of sets satisfying the desired conditions. To see this, we need to show (1) $\mathcal{F} \setminus \mathcal{S}$ is p -covered and (2) $|\mathcal{S}| \leq 2^H$.

Step 1: $\mathcal{F} \setminus \mathcal{S}$ is p -covered. If $\mathcal{F} \setminus \mathcal{S} = \emptyset$, this is trivially true. Otherwise, suppose for contradiction that $\mathcal{F} \setminus \mathcal{S}$ is not p -covered. Then Proposition 4.8(2) yields a p -smooth distribution μ supported on $\mathcal{F} \setminus \mathcal{S}$. For $\varepsilon \in (0, 1)$, define $\nu_\varepsilon := (1 - \varepsilon)\nu^* + \varepsilon\mu$. Since this is a convex combination of p -smooth distributions over $\mathcal{F}$, we have $\nu_\varepsilon \in \mathcal{P}$. Now,

- For any $S \in \mathcal{S}$, $\mu(S) = 0$ as μ is supported on $\mathcal{F} \setminus \mathcal{S}$. Therefore, $\nu_\varepsilon(S) = (1 - \varepsilon)\tau^* < \tau^*$.
- For any $T \in \mathcal{F} \setminus \mathcal{S}$, we have $\nu^*(T) < \tau^*$ by the definition of $\mathcal{S}$. Since $\mathcal{F}$ is finite, the minimum gap $\delta := \min_{T \in \mathcal{F} \setminus \mathcal{S}} (\tau^* - \nu^*(T))$ is strictly positive. As $\mu(T) \leq 1$ for all T , we find

$$\nu_\varepsilon(T) = (1 - \varepsilon)\nu^*(T) + \varepsilon\mu(T) = \nu^*(T) + \varepsilon(\mu(T) - \nu^*(T)) \leq \nu^*(T) + \varepsilon.$$

Therefore if we choose $\varepsilon \in (0, \delta)$, we get

$$\nu_\varepsilon(T) \leq \nu^*(T) + \varepsilon < \nu^*(T) + \delta \leq \tau^*$$

for all $T \in \mathcal{F} \setminus \mathcal{S}$, contradicting the optimality of ν^* . Hence, $\mathcal{F} \setminus \mathcal{S}$ must be p -covered.

Step 2: $|\mathcal{S}| \leq 2^H$. Since ν^* assigns mass τ^* to each $S \in \mathcal{S}$, we have $|\mathcal{S}|\tau^* \leq 1$ which implies $|\mathcal{S}| \leq 1/\tau^*$. On the other hand,

$$H(\nu^*) = \sum_{T \in \mathcal{F}} \nu^*(T) \log \frac{1}{\nu^*(T)} \geq \sum_{T \in \mathcal{F}} \nu^*(T) \log \frac{1}{\|\nu^*\|_\infty} = \log \frac{1}{\tau^*}.$$

Since by assumption $H(\nu^*) \leq H$, this implies $\tau^* \geq 2^{-H}$. As a result, $|\mathcal{S}| \leq 2^H$. $\square$

4.4 Universe reduction

We are now ready to prove Theorem 1.8. We start by making the following definition:

Definition 4.9 ((p, M) -almost-covered). *Let $\mathcal{F} \subseteq 2^{[n]}$. We say $\mathcal{F}$ is (p, M) -almost-covered if for every $J \subseteq [n]$, there exists an exceptional set $\mathcal{S}_J \subseteq \mathcal{F}_J$ of size $|\mathcal{S}_J| \leq M$ such that $\mathcal{F}_J \setminus \mathcal{S}_J$ is p -covered.*

We show that if a set family $\mathcal{F}$ is almost-covered, then we can identify a small set of coordinates I that covers most of the sets in $\mathcal{F}$.

Lemma 4.10. *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w -sets. Fix $p \in (0, 1)$, $M \geq 1$ and $\delta \in (0, 1]$. If $\mathcal{F}$ is (p, M) -almost-covered, then there exists an $I \subseteq [n]$ with $|I| \leq t$ such that*

$$|\mathcal{F}_I| \leq \delta |\mathcal{F}| + tM.$$

where $t := \lceil \frac{2}{p}(w \ln 2) + \ln(1/\delta) \rceil$.

In words, the lemma is saying that there exists a small universe of size t on which most sets in $\mathcal{F}$ are preserved.

Proof. Without loss of generality, we can assume $\mathcal{F}$ is non-empty. Otherwise, we can take $I = \emptyset$ and the conclusion follows trivially. Since $\mathcal{F}$ is (p, M) -almost-covered by assumption, we can fix a choice of witness $(\mathcal{S}_J, Q_J)$ for each $J \subseteq [n]$ as in Definition 4.9. That is, $\mathcal{S}_J \subset \mathcal{F}_J$ of size $|\mathcal{S}_J| \leq M$, and Q_J is a distribution over J that p -covers $\mathcal{F}_J \setminus \mathcal{S}_J$. We now describe a process that iteratively selects a set of important elements that cover most sets in $\mathcal{F}$. Let t_{end} be some stopping condition to be determined later. Then:

- Initialize $I_0 := \emptyset$, $J_0 = [n]$, $\mathcal{S}_{\text{removed}}^{(0)} = \emptyset$ and $\mathcal{F}_0 = \mathcal{F}$.
- For $j = 0, 1, \dots, t_{\text{end}} - 1$:
 - Let $(\mathcal{S}_{J_j}, Q_{J_j})$ denote the corresponding witness for J_j .
 - Remove exceptions: $\mathcal{F}_j^- := \mathcal{F}_j \setminus \mathcal{S}_{J_j}$.
 - Sample $i_{j+1} \sim Q_{J_j}$ and add it: $I_{j+1} = I_j \cup \{i_{j+1}\}$.
 - Update $J_{j+1} := [n] \setminus I_{j+1}$ and $\mathcal{S}_{\text{removed}}^{(j+1)} = \mathcal{S}_{\text{removed}}^{(j)} \cup \mathcal{S}_{J_j}$.
 - Set $\mathcal{F}_{j+1} = (\mathcal{F}_j^-)_{J_{j+1}}$, and remove any empty sets or duplicate sets if some exist.

Now define the potential function⁴:

$$\Phi_j := \sum_{A \in \mathcal{F}_j} 2^{|A \cap J_j|}.$$

It is easy to see that Φ_j is decreasing in j and $\Phi_j \geq 2|\mathcal{F}_j|$. Moreover, since $\mathcal{F}$ is a family of w -sets, we know $\Phi_0 \leq |\mathcal{F}| \cdot 2^w$. Let t_{end} be the smallest j such that $\mathbb{E}[\Phi_j] \leq \delta |\mathcal{F}|$. We claim $t_{\text{end}} := \min\{n, \lceil \frac{2}{p}(w \ln 2) + \ln(1/\delta) \rceil\}$.

To this end, we first observe that the process can continue at iteration j as long as $|\mathcal{F}_j| \geq M$. Lets assume this is always the case before we finish (we will justify this assumption later). Now condition on the choice of $i_1, \dots, i_j$ so that $I_j, J_j, \mathcal{F}_j$ are fixed. Let $A \in \mathcal{F}_j^-$. Then we have $A \notin \mathcal{S}_{J_j}$, which implies $\Pr_{i_{j+1} \sim Q_{J_j}}[i_{j+1} \in A] \geq p$. Moreover, we know

$$|A \cap J_{j+1}| = |A \cap J_j| - \mathbf{1}[i_{j+1} \in A].$$

Therefore,

$$\mathbb{E}_{i_{j+1} \sim Q_{J_j}} \left[2^{|A \cap J_{j+1}|} \mid i_1, \dots, i_j \right] = \mathbb{E}[2^{|A \cap J_j| - \mathbf{1}[i_{j+1} \in A]}] = 2^{|A \cap J_j|} \cdot \mathbb{E}[2^{-\mathbf{1}[i_{j+1} \in A]}] \leq 2^{|A \cap J_j|} \cdot \left(1 - \frac{p}{2}\right).$$

⁴Since $A \in \mathcal{F}_j$, we have $|A \cap J_j| = |A|$. Here we write $|A \cap J_j|$ to emphasize it is the size of an intersection.

By linearity of expectation, we thus find

$$\begin{aligned}
\mathbb{E}_{i_{j+1} \sim Q_{J_j}} [\Phi_{j+1} \mid i_1, \dots, i_j] &= \sum_{B \in \mathcal{F}_{j+1}} \mathbb{E}_{i_{j+1} \sim Q_{J_j}} [2^{|B \cap J_{j+1}|} \mid i_1, \dots, i_j] \\
&\leq \sum_{A \in \mathcal{F}_j^-} \mathbb{E}_{i_{j+1} \sim Q_{J_j}} [2^{|A \cap J_{j+1}|} \mid i_1, \dots, i_j] \\
&\leq \sum_{A \in \mathcal{F}_j^-} 2^{|A \cap J_j|} \cdot \left(1 - \frac{p}{2}\right) \\
&\leq \sum_{A \in \mathcal{F}_j} 2^{|A \cap J_j|} \cdot \left(1 - \frac{p}{2}\right) = \left(1 - \frac{p}{2}\right) \Phi_j,
\end{aligned}$$

where the first inequality holds since for every $B \in \mathcal{F}_{j+1}$, there exists some $A \in \mathcal{F}_j^-$ such that $B \subseteq A$. The last inequality holds since $\mathcal{F}_j^-$ is a subset of $\mathcal{F}_j$. Iterating and using $1 - x \leq e^{-x}$, we obtain

$$\mathbb{E}_{i_1, \dots, i_t} [\Phi_t] \leq \left(1 - \frac{p}{2}\right)^t \Phi_0 \leq \exp(-pt/2) \cdot (|\mathcal{F}| \cdot 2^w).$$

Setting $\delta = 2^w \cdot \exp(-pt_{\text{end}}/2)$ and solving for t_{end} , we get $t_{\text{end}} = \lceil \frac{2}{p}(w \ln 2) + \ln(1/\delta) \rceil$. Now by the first moment method, we know there exists a sequence of $I = \{i_1, \dots, i_{t_{\text{end}}}\}$ such that $\sum_{A \in \mathcal{F}_{t_{\text{end}}}} 2^{|A|} \leq \delta |\mathcal{F}|$. Therefore, the number of sets in $\mathcal{F}$ that are not covered by I is

$$|\mathcal{F}_J| \leq |\mathcal{F}_{t_{\text{end}}}| + |\mathcal{S}_{\text{removed}}^{t_{\text{end}}}| \leq \delta |\mathcal{F}| + t_{\text{end}} M.$$

The desired result then follows since I can have size at most n . Lastly, if during any iteration before termination we encounter the case where $|\mathcal{F}_j| \leq M$ which clearly satisfies the statement. Hence, our assumption $|\mathcal{F}_j| \geq M$ for all iterations is a valid one. $\square$

We have thus seen (p, M) -almost-coveredness is a desirable property as it allows us to reduce the size of the universe without affecting the size of the set family too much. It remains to understand how we can get such a nice property. It turns out we can obtain almost-coveredness from moonflower-freeness.

Lemma 4.11 (Almost-coveredness from moonflower-freeness). *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k -moonflower-free family of w -sets. Then $\mathcal{F}$ is $(p, 2^h)$ -almost-covered where*

$$h = h(w, k, p) := O\left(p \log(1/p) \log \binom{k + w \lceil 2/p \rceil}{k}\right).$$

Proof. Since $\mathcal{F}$ is k -moonflower-free, by Lemma 2.4, each projection $\mathcal{F}_J$ is also k -moonflower-free. By Theorem 4.6, every p -smooth distribution D supported on $\mathcal{F}_J$ must satisfy

$$H(D) \leq O\left(p \log(1/p) \log \binom{k + w \lceil 2/p \rceil}{k}\right).$$

Now by Theorem 4.7, for each J there exists an exceptional set $\mathcal{S}_J \subseteq \mathcal{F}_J$ with $|\mathcal{S}_J| \leq 2^h$ such that $\mathcal{F}_J \setminus \mathcal{S}_J$ is p -covered. In particular, this implies $\mathcal{F}$ is $(p, 2^h)$ -almost-covered. $\square$

We are now ready to prove Theorem 1.8, restated below for convenience.

Theorem 4.12 (Universe reduction, Theorem 1.8 restated). *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k -moonflower-free family of w -sets. Then for any $\lambda \in (0, 1]$, there exists $I \subseteq [n]$ of size $|I| \leq O(\lambda^{-1} k L^3)$ satisfying $|\mathcal{F}_{\bar{I}}| \leq \exp(\lambda w)$ where $L := \log(kw/\lambda)$.*

Proof. Assume without loss of generality that $|\mathcal{F}| > \exp(\lambda w)$. Otherwise, take $I = \emptyset$ and we are done. Let $p := c\lambda w/(kL^2)$ where c is some sufficiently small constant, and; we have⁵ $p < 1/2$. Let $r = \lceil 2/p \rceil$. By Lemma 4.11, we know $\mathcal{F}$ is $(p, 2^h)$ -almost-covered with $h := O(p \log(1/p) \cdot \log \binom{k+w}{k})$. For our choice of p , we have $\log(1/p) = O(L)$ and

$$\log \binom{k+w}{k} \leq k \log \left(\frac{e(k+w)}{k} \right) \leq O(kL).$$

Consequently, $h \leq O(c\lambda w) \leq \lambda w/2$ for c a small enough constant. Applying Lemma 4.10 with $p := c\lambda w/(kL^2)$, $M = 2^h$ and $\delta = 1/|\mathcal{F}|$, we get that there exists $I \subseteq [n]$ such that $|I| \leq t := \lceil \frac{2}{p} \log(|\mathcal{F}|2^h) \rceil + \log |\mathcal{F}|$ and $|\mathcal{F}_{\bar{I}}| \leq t \cdot 2^h$. By the extremal moonflower bound (Theorem 3.1), we get

$$t \leq O \left(\frac{\log |\mathcal{F}| + w}{p} + \log |\mathcal{F}| \right) \leq O \left(\frac{\log \binom{k+w}{w} + w}{p} \right) \leq O \left(\frac{wL}{p} \right) \leq O \left(\frac{kL^3}{\lambda} \right),$$

where the last inequality holds by definition of p . If $t < 2^{\lambda w/2}$, then $|\mathcal{F}_{\bar{I}}| \leq t \cdot 2^{\lambda w/2} \leq \exp(\lambda w)$ and we are done.

Otherwise, we have $\lambda w \leq 2 \log t$. In this case we take $I := \text{supp}(\mathcal{F})$. Then $\mathcal{F}_{\bar{I}} = \{\emptyset\}$. Moreover, by Lemma 2.6, we know

$$|I| \leq kw \leq \frac{2k \log t}{\lambda} \leq O \left(\frac{kL}{\lambda} \right)$$

which is even better than the claimed bound. $\square$

4.5 Universe reduction lower bounds

In this subsection, we prove lower bounds on the size of punctured sets needed for universe reduction. We first justify why the interesting regime of parameters is $k \gg \lambda w \gg 1$:

1. If $k = O(\lambda w)$ then by the extremal moonflower bound, $|\mathcal{F}| \leq \binom{k+w-1}{k} \leq \exp(O(\lambda w \log(1/\lambda)))$ which is already close enough to the desired trace bound $\exp(\lambda w)$ by slightly adjusting λ .
2. If $\lambda = O(1/w)$ then the desired trace bound $\exp(\lambda w)$ is $O(1)$; and the support of $\mathcal{F}$ is $|\text{supp}(\mathcal{F})| \leq (k-1)w \leq O(k/\lambda)$. So we can simply prune all elements in the support.

Lemma 4.13 (Lemma 1.9, restated). *Let $k, w \geq 2$ be positive integers and $\lambda \in (0, 1]$. Assume that $k = \Omega(\lambda w)$ and $\lambda w = \Omega(1)$. Then there exists a family of w -sets $\mathcal{F} \subseteq 2^{[n]}$ which is k -moonflower-free such that any subset $I \subseteq [n]$ satisfying $|\mathcal{F}_{\bar{I}}| \leq \exp(\lambda w)$ must have size $|I| = \Omega(\lambda^{-1}k)$.*

Proof. Define $r := 1/\lambda$ and $m := \lambda w$, and assume without loss of generality that r, m are integers, by changing λ, w by constant factors if needed. Let $p := k + m - 2$ and $\mathcal{G}$ be the family of all subsets of $[p]$ of size m . We saw in Lemma 3.3 that $\mathcal{G}$ is k -moonflower free. Next, consider the set family $\mathcal{F}$ defined as $\mathcal{F} := \{A \times [r] : A \in \mathcal{G}\} \subseteq 2^{[p] \times [r]}$, and note that all sets in $\mathcal{F}$ have size $mr \leq w$. We claim that $\mathcal{F}$ is also k -moonflower free. Indeed, if $S_1, \dots, S_k$ form a k -moonflower in $\mathcal{F}$, then by construction $S_i = A_i \times [r]$ and $A_1, \dots, A_k \in \mathcal{G}$ also form a k -moonflower, a contradiction.

Next, consider an arbitrary set $I \subseteq [p] \times [r]$. For each $a \in [r]$, define $I_a := \{x \in [p] : (x, a) \in I\}$. Notice that $|\mathcal{F}_{\bar{I}}| \geq |\mathcal{G}_{\bar{I}_a}|$ for all $a \in [r]$. Hence, if $|\mathcal{F}_{\bar{I}}| \leq \exp(m)$, then we necessarily have $|\mathcal{G}_{\bar{I}_a}| \leq \exp(m)$ for all $a \in [r]$.

On the other hand, let $A \subseteq [p]$. Then by the definition of $\mathcal{G}$, $\mathcal{G}_{\bar{A}}$ contains all subsets of size m from the remaining $p - |A|$ coordinates. Hence, $|\mathcal{G}_{\bar{A}}| \geq \binom{p-|A|}{m} \geq \left(\frac{p-|A|}{m} \right)^m \geq \exp(m)$ provided $p - |A| \geq em$. Therefore, if $|\mathcal{G}_{\bar{A}}| \leq \exp(m)$, we must have $p - |A| < em$. Equivalently, $|A| > k - (e-1)m - 2$.

Applying this with $A = I_a$, we know $|I_a| \geq k - (e-1)m - 2$ for all $a \in [r]$. As a result,

$$|I| = \sum_{a \in [r]} |I_a| \geq r(k - (e-1)m - 2) \geq \frac{k}{\lambda} - O(w),$$

⁵This follows as $p \leq \frac{c \log |\mathcal{F}|}{k \log^2(kw/\lambda)} \leq \frac{c \log \binom{k+w}{k}}{k \log^2(kw/\lambda)} \leq \frac{ck \log(e(1+w/k))}{k \log^2(kw/\lambda)} < 1/2$ for c small enough.

where the last inequality follows by our choice of r and m . As we assume $k = \Omega(\lambda w)$, we conclude that $|I| = \Omega(k/\lambda)$. $\square$

5 Improved sparsification with a single $\log n$ factor

In this section, we refine the sparsification analysis from [BG25]. The only dependence on the ambient blocklength n is the *single* $\log n$ factor coming from the recursion depth. All additional losses are $\text{poly}(\log(k/\varepsilon), \log \log n)$.

Throughout, let $\mathcal{C} \subseteq \{0, 1\}^n$ be a code with $\text{NRD}(\mathcal{C}) \leq k-1$. Equivalently, the support family $\text{Supp}(\mathcal{C}) := \{\text{supp}(x) : x \in \mathcal{C}\}$ is k -moonflower-free by Lemma 2.8.

A *weighted coordinate set* is a pair (T, α) where $T \subseteq [n]$ and $\alpha : T \rightarrow \mathbb{R}_{\geq 0}$. It induces the estimator

$$\widehat{\text{wt}}_{T, \alpha}(x) := \sum_{i \in T} \alpha(i) x_i, \quad x \in \{0, 1\}^n.$$

We say (T, α) ε -*sparsifies* $\mathcal{C}$ if $\widehat{\text{wt}}_{T, \alpha}(x) \in (1 \pm \varepsilon) \text{wt}(x)$ for all $x \in \mathcal{C}$ where $\text{wt}(x) = \sum_{i=1}^n x_i$. Such a pair (T, α) is called an ε -*sparsifier* of $\mathcal{C}$. Our main sparsification result is as follows.

Theorem 5.1 (Improved sparsification with a single $\log n$ factor). *Let $\mathcal{C} \subseteq \{0, 1\}^n$ satisfy $\text{NRD}(\mathcal{C}) \leq k-1$. Then for every $\varepsilon \in (0, 1/4)$ there exists a weighted coordinate set (T, α) that ε -sparsifies $\mathcal{C}$ and satisfies*

$$|T| \leq \frac{k \log n}{\varepsilon^2} \cdot \text{poly}(\log(k/\varepsilon), \log \log n).$$

5.1 Proof of Theorem 5.1

The proof of Theorem 5.1 is a recursive “puncture-then-halve” process run for $R = \lceil \log_2 n \rceil$ rounds. The main calculations to keep in mind are:

- Each round adds a puncturing set I_r of size at most $\frac{k}{\varepsilon^2} \cdot \text{poly}(\log(k/\varepsilon), \log \log n)$. Since there are $R = \Theta(\log n)$ rounds, the final puncturing set has size $|T| \leq \frac{k \log n}{\varepsilon^2} \cdot \text{poly}(\log(k/\varepsilon), \log \log n)$. The remaining coordinates will be small due to halving in each round.
- The puncturing is carried-out carefully in that for each weight range w , we choose a specific puncturing set $I_{r,w}$. For round r , let $\mathcal{F}_{(w, 2w]}^{(r)}$ denote codewords whose weight in round r is between $w, 2w$ (in the non-punctured portion).

We ensure that the number of traces of this family outside $I_{r,w}$ is roughly $\exp(\eta_0^2 w)$ where $\eta_0 \approx \varepsilon$. This makes the union bound combined with Chernoff viable for the codewords whose weights are in $[w, 2w]$: with good probability, randomly sampling half the remaining coordinates preserves the weight of all such codewords within a factor of $(1 \pm \eta_0)$. The one caveat is that we need the failure probability to be at most $1/\text{poly}(\log n)$ so we have a minimum weight threshold $w_{\min}$ above which we use this argument.

- The above argument is still problematic as each round incurs a multiplicative error of $(1 + \eta_0)$ and we have $O(\log n)$ rounds. The simple fix of taking $\eta_0 \ll \varepsilon/(\log n)$ would in turn blow up the support size to be $\text{poly}(\log n)$ which we want to avoid.

The key point is that the error in the Chernoff bound combined with union bound argument can be better for large weight codewords. We set a weight-threshold $w_{\text{high}} \approx O((k \log(k/\varepsilon) + \log \log n)/\varepsilon^2)$ and do the following. Pick a dyadic weight $w = 2^i w_{\text{low}} > w_{\text{high}}$. For such codewords, we do no puncturing, and instead union-bound directly over such codewords using our optimal moonflower bound. Essentially, the number of codewords whose weight is between $(w, 2w]$ is at most $(Cw/k)^k$. This allows to say that with high probability, randomly picking half the remaining codewords preserves the weights up to error roughly

$$\eta(w) = O(\sqrt{(k \log(w/k) + \log \log n)/w}).$$

Note that the error decreases as w increases.

- **Error accumulation:** Finally, we do an error analysis for each codeword conditioned on the union bounds succeeding across all rounds. Fix a codeword $x \in \mathcal{C}$ and let w_r denote its weight among non-punctured coordinates in round r . Then, the multiplicative error, $1 \pm \varepsilon_r$, incurred by the sampling process satisfies

$$\varepsilon_r := \begin{cases} 0 & w_r \leq w_{\text{low}}, \\ \eta_0 & w_{\text{low}} < w_r \leq w_{\text{high}} \\ \eta(w_r) & w_r > w_{\text{high}} \end{cases}$$

In each round, the residual weight of a codeword drops by a constant factor as long as its weight is above w_{low} . Thus, it crosses the medium region $(w_{\text{low}}, w_{\text{high}}]$ in only $O(\log w_{\text{high}})$ rounds and the total contribution to the multiplicative error through these rounds is $O(\eta_0 \log w_{\text{high}})$, and we set $\eta_0 = \Theta(\varepsilon / \log w_{\text{high}})$.

There can be several rounds where the codeword would be large weight (i.e, weight above w_{high}). However, the error for these rounds is better, and we get a convergent geometric series for the errors here because of the improved bound on $\eta(w)$ above.

Proof. (of Theorem 5.1)

Parameter choices. Fix $R := \lceil \log_2 n \rceil$. In what follows, let C be a large constant. Define the transition threshold

$$w_{\text{high}} := \left\lceil C \cdot \frac{k \log(k/\varepsilon) + \log R}{\varepsilon^2} \right\rceil.$$

Define the medium-weight per-round error

$$\eta_0 := \frac{\varepsilon}{100 \log(2w_{\text{high}})}.$$

Define the low-weight cutoff

$$w_{\text{low}} := \left\lceil \frac{C}{\eta_0^2} \cdot (\log(k/\varepsilon) + \log R) \right\rceil.$$

Finally, for dyadic $w > w_{\text{high}}$, define the large-weight per-round error

$$\eta(w) := \min \left\{ \frac{1}{4}, \sqrt{\frac{C(k \log(w/k) + \log R)}{w}} \right\}.$$

Recursive construction. Let $U_0 = [n]$. For rounds $r = 0, 1, \dots, R-1$, given U_r we choose a puncturing set $I_r \subseteq U_r$, set $V_r := U_r \setminus I_r$, and then form U_{r+1} by including each element of V_r independently with probability $1/2$.

For dyadic⁶ $w \in [w_{\text{low}}, w_{\text{high}}]$, define

$$\mathcal{F}_{(w, 2w)}^{(r)} := \{ \text{supp}(x) \cap U_r : x \in \mathcal{C}, w < |\text{supp}(x) \cap U_r| \leq 2w \}.$$

Define the low-weight family

$$\mathcal{F}_{\leq w_{\text{low}}}^{(r)} := \{ \text{supp}(x) \cap U_r : x \in \mathcal{C}, |\text{supp}(x) \cap U_r| \leq w_{\text{low}} \}.$$

(All these families are k -moonflower-free by Lemma 2.4.)

(a) *Low weights:* capture deterministically. Let

$$I_{r, \text{low}} := \text{supp} \left(\mathcal{F}_{\leq w_{\text{low}}}^{(r)} \right) \subseteq U_r.$$

By Lemma 2.6, $|I_{r, \text{low}}| \leq k w_{\text{low}}$.

⁶Here by dyadic, we mean $w = 2^j w_{\text{low}}$ for $j = 0, 1, \dots, \log(w_{\text{high}}/w_{\text{low}})$.

- (b) *Medium weights:* puncture to shrink traces. Choose a sufficiently small absolute constant $\theta_0 > 0$, say $\theta_0 = 10^{-3}$ after adjusting constants in the Chernoff bound. For each dyadic $w \in [w_{\text{low}}, w_{\text{high}}]$, apply Theorem 1.8 to $\mathcal{F}_{(w, 2w]}^{(r)}$ with parameter $\lambda = \theta_0 \eta_0^2$ to obtain $I_{r,w} \subseteq U_r$ such that

$$\left| \left(\mathcal{F}_{(w, 2w]}^{(r)} \right)_{\overline{I_{r,w}}} \right| \leq \exp(\theta_0 \eta_0^2 w), \quad (1)$$

and

$$|I_{r,w}| \leq \frac{k}{\eta_0^2} \cdot \text{poly} \left(\log \frac{kw}{\eta_0} \right).$$

Here we use the fact that $\eta_0 < 1$ and θ_0 is an absolute constant, so its contribution is absorbed into the polylogarithmic factor.

- (c) *Define I_r and recurse.* Let

$$I_r := I_{r,\text{low}} \cup \bigcup_{\substack{\text{dyadic } w \\ w_{\text{low}} \leq w \leq w_{\text{high}}}} I_{r,w}, \quad V_r := U_r \setminus I_r,$$

and sample $U_{r+1} \subseteq V_r$ by keeping each coordinate with probability $1/2$.

- (d) *Output.* Define

$$T := \left(\bigcup_{r=0}^{R-1} I_r \right) \cup U_R, \quad \alpha(i) := 2^r \text{ if } i \in I_r, \quad \alpha(i) := 2^R \text{ if } i \in U_R.$$

(The sets $I_0, \dots, I_{R-1}, U_R$ are disjoint since $U_{r+1} \subseteq U_r \setminus I_r$.)

Successful Sampling. Fix a round r and condition on the entire history up to round r so that U_r, I_r, V_r are fixed.

Medium regime. Fix dyadic $w \in [w_{\text{low}}, w_{\text{high}}]$. Consider the trace family on V_r :

$$\left(\mathcal{F}_{(w, 2w]}^{(r)} \right)_{V_r}.$$

Since $V_r \subseteq U_r \setminus I_{r,w}$, restriction cannot increase size, so

$$\left| \left(\mathcal{F}_{(w, 2w]}^{(r)} \right)_{V_r} \right| \leq \left| \left(\mathcal{F}_{(w, 2w]}^{(r)} \right)_{\overline{I_{r,w}}} \right| \leq \exp(\theta_0 \eta_0^2 w).$$

Moreover, for any A in this family, we have $|A| \leq 2w$. Applying Lemma 2.12 with $\Delta := \eta_0 w$ gives

$$\mathbf{Pr}_{U_{r+1}} \left[\left| 2|A \cap U_{r+1}| - |A| \right| > \eta_0 w \right] \leq 2 \exp(-\Omega(\eta_0^2 w)).$$

Using (1) and the definition of w_{low} , the union bound over all $A \in (\mathcal{F}_{(w, 2w]}^{(r)})_{V_r}$ fails with probability at most $\exp(-\Omega(\eta_0^2 w)) \leq 1/(100R \log(2w_{\text{high}}))$ for appropriate choice of constants C, θ_0 .

Large regime. Fix dyadic $w > w_{\text{high}}$ and consider the trace family

$$\mathcal{H}_{(w, 2w]}^{(r)} := \left(\mathcal{F}_{(w, 2w]}^{(r)} \right)_{V_r}.$$

Since $\mathcal{F}_{(w, 2w]}^{(r)}$ is k -moonflower-free and $w \geq k$ in this regime, Theorem 1.4 gives

$$|\mathcal{H}_{(w, 2w]}^{(r)}| \leq |\mathcal{F}_{(w, 2w]}^{(r)}| \leq \left(\frac{C_0 w}{k} \right)^k,$$

for some constant $C_0 > 0$. For any $A \in \mathcal{H}_{(w, 2w)}^{(r)}$ we have $|A| \leq 2w$, so by Lemma 2.12 with $\Delta := \eta(w)w$,

$$\Pr_{U_{r+1}} [|2|A \cap U_{r+1}| - |A| | > \eta(w)w] \leq 2 \exp(-\Omega(\eta(w)^2 w)) = 2 \exp(-\Omega(k \log(ew/k) + \log R)).$$

For large enough C in the definition of $\eta(w)$, the union bound over all $A \in \mathcal{H}_{(w, 2w)}^{(r)}$ fails with probability at most $1/(100R \cdot 2^{j+2})$ when $w = 2^j$. Summing over dyadic $w > w_{\text{high}}$ gives failure probability at most $1/(100R)$.

Combining medium and large regimes and summing over the $O(\log w_{\text{high}})$ medium scales, we obtain that conditioned on the past, round r fails with probability at most $1/(50R)$. A union bound over $r = 0, \dots, R-1$ yields that all rounds succeed with probability at least $49/50$. Also $\mathbb{E}[|U_R|] \leq n2^{-R} \leq 1$, hence $\Pr[|U_R| \leq 50] \geq 49/50$ by Markov. Intersecting gives overall success probability at least $2/3$.

Accuracy for a fixed codeword. Fix a successful outcome as above. Fix a codeword $x \in \mathcal{C}$ with $S := \text{supp}(x)$.

Let $\hat{N}_r = \sum_{i=0}^{r-1} 2^i |S \cap I_i| + 2^r |S \cap U_r|$ denote the weight estimate for the weight of x at round r . Note that $\hat{N}_0 = |S|$ is the Hamming weight of x . Our goal is to show that $\hat{N}_R = (1 \pm \varepsilon)\hat{N}_0$.

Consider the next weight estimate

$$\hat{N}_{r+1} = \sum_{i=0}^{r-1} 2^i |S \cap I_i| + 2^r (|S \cap I_r| + 2|S \cap U_{r+1}|).$$

Let $w_r := |S \cap U_r|$ be the true residual weight at round r , and let

$$\varepsilon_r := \begin{cases} 0 & w_r \leq w_{\text{low}}, \\ \eta_0 & w_{\text{low}} < w_r \leq w_{\text{high}}, \\ \eta(w_r) & w_r > w_{\text{high}}. \end{cases}$$

Then, as we are in a successful outcome, we have,

$$|2|S \cap U_{r+1}| - |S \cap (U_r \setminus I_r)| \leq \varepsilon_r w_r.$$

Thus,

$$2|S \cap U_{r+1}| + |S \cap I_r| - |S \cap U_r| \leq \varepsilon_r w_r.$$

Combining the above equations, we get

$$|\hat{N}_{r+1} - \hat{N}_r| = 2^r \cdot |2|S \cap U_{r+1}| + |S \cap I_r| - |S \cap U_r| \leq \varepsilon_r 2^r w_r \leq \varepsilon_r \hat{N}_r.$$

Thus, we have

$$(1 - \varepsilon_r)\hat{N}_r \leq \hat{N}_{r+1} \leq (1 + \varepsilon_r)\hat{N}_r.$$

Applying the above inequality for $r = 0, \dots, R-1$, we get

$$\hat{N}_0 \cdot \prod_{r=0}^{R-1} (1 - \varepsilon_r) \leq \hat{N}_R \leq \hat{N}_0 \cdot \prod_{r=0}^{R-1} (1 + \varepsilon_r).$$

Further, as $\varepsilon_r < 1/2$, the above can be simplified to

$$\hat{N}_0 \cdot \exp\left(-2 \sum_{r=0}^{R-1} \varepsilon_r\right) \leq \hat{N}_R \leq \hat{N}_0 \cdot \exp\left(\sum_{r=0}^{R-1} \varepsilon_r\right). \quad (2)$$

It remains to bound $\sum_r \varepsilon_r$. The number of rounds with $w_{\text{low}} < w_r \leq w_{\text{high}}$ is $O(\log(2w_{\text{high}}))$ since $w_{r+1} \leq (1 + \eta_0)w_r/2 \leq (3/5)w_r$. Thus $\sum_{w_{\text{low}} < w_r \leq w_{\text{high}}} \varepsilon_r \leq O(\eta_0 \log(2w_{\text{high}})) \leq \varepsilon/50$.

For rounds with $w_r > w_{\text{high}}$, we have $\varepsilon_r \leq \eta(w_r)$ and $w_{r+1} \leq (1 + \varepsilon_r)w_r/2 \leq (5/8)w_r$. A geometric-series estimate gives

$$\sum_{w_r > w_{\text{high}}} \varepsilon_r \leq O\left(\sqrt{\frac{k \log(ew_{\text{high}}/k) + \log R}{w_{\text{high}}}}\right) \leq \varepsilon/50$$

by the definition of w_{high} (with C large enough).

Combining the above we get that $\sum_r \varepsilon_r \leq \varepsilon/25$. Thus, in particular we must have $\hat{N}_R = (1 \pm \varepsilon)N_0$, as we wanted.

Size bound. We have $|T| \leq \sum_{r=0}^{R-1} |I_r| + |U_R|$ and $|U_R| \leq 50$ on the good event. Also $|I_{r,\text{low}}| \leq kw_{\text{low}}$.

For each dyadic $w \in [w_{\text{low}}, w_{\text{high}}]$, apply Theorem 1.8 to the layer family $\mathcal{F}_{(w,2w)}^{(r)}$ with $\lambda = \theta_0 \eta_0^2$ to obtain

$$|I_{r,w}| \leq \frac{k}{\eta_0^2} \cdot \text{poly}(\log(kw/\eta_0)).$$

Since $w \leq w_{\text{high}}$ throughout, we have $\log w \leq \log w_{\text{high}}$ and

$$\log(k/\eta_0) = \log(k/\varepsilon) + O(\log \log w_{\text{high}}),$$

hence

$$|I_{r,w}| \leq \frac{k}{\eta_0^2} \cdot \text{poly}(\log(k/\varepsilon), \log \log n),$$

using $\log w_{\text{high}} = O(\log(k/\varepsilon) + \log \log n)$ by the definition of w_{high} .

There are $O(\log(2w_{\text{high}}))$ such dyadic values. Thus

$$|I_r| \leq \frac{k}{\eta_0^2} \cdot \text{poly}(\log(k/\varepsilon), \log \log n),$$

since $\eta_0^{-2} = \Theta(\log^2(2w_{\text{high}})/\varepsilon^2)$ and $w_{\text{low}} = \Theta(\eta_0^{-2}(\log(k/\varepsilon) + \log R))$. Finally $R = \Theta(\log n)$ gives

$$|T| \leq \frac{k \log n}{\varepsilon^2} \cdot \text{poly}(\log(k/\varepsilon), \log \log n),$$

as claimed.

This finishes the proof of the full sparsification theorem. $\square$

5.2 Lower bound

In this subsection, we state and prove our lower bound construction.

Lemma 5.2 (Lemma 1.12, restated). *Let $k \geq 1$ and $\varepsilon \in (0, 1)$. Then, for all large enough n , there exists an explicit $\mathcal{C} \subseteq \{0, 1\}^n$ with $\text{NRD}(\mathcal{C}) = k$ such that any ε -sparsifier (T, α) of $\mathcal{C}$ must satisfy*

$$|T| = \Omega\left(\frac{k \log(n/k)}{\varepsilon}\right).$$

Proof. We will assume n is chosen large enough so that $\frac{k \log(n/k)}{\varepsilon} = O(n)$. In the construction we identify $\{0, 1\}^n$ with subsets of $[n]$. Assume without loss of generality that k divides n and let $m = n/k$. Let $a_1 \leq \dots \leq a_s$ be a maximal collection of integers such that $1 \leq a_1$, $a_s < m$ and $a_{j+1} > (1 + \varepsilon)a_j/(1 - \varepsilon)$ for all $j < s$. Clearly, such numbers exist for $s = \Omega(\log(n/k)/\varepsilon)$ (here is where we use the assumption that n is large enough, as we clearly have $s \leq n/k$). For two integers $i < j$ let $[i : j] = \{i, i+1, \dots, j\}$. For $i \in [k], j \in [s]$ define the set

$$S_{i,j} = [(i-1)m : (i-1)m + a_j].$$

We take $\mathcal{C}_i := \{S_{i,j} : j \in [s]\}$ and $\mathcal{C} := \cup_{i \in [k]} \mathcal{C}_i$.

Observe that $\mathcal{C}_i$ are defined on disjoint ground sets and that each $\mathcal{C}_i$ is a chain. This implies that $\text{NRD}(\mathcal{C}_i) = 1$ and hence ⁷ $\text{NRD}(\mathcal{C}) = k$. Finally, let (T, α) be an ε -sparsifier for $\mathcal{C}$. We claim that $|T| \geq |\mathcal{C}|$ which concludes the proof.

Assume this is not the case. Then there must exist some $i \in [k]$ such that $|T \cap \text{Supp}(\mathcal{C}_i)| < |\mathcal{C}|/k = s$. In particular, there must exist $i \in [k]$ and $j < s$ such that

$$T \cap S_{i,j} = T \cap S_{i,j+1}.$$

This however cannot be the case as by construction $|S_{i,j+1}| > (1 + \varepsilon)|S_{i,j}|$ and (T, α) is an ε -sparsifier for $\mathcal{C}$. $\square$

⁷A size $k \times k$ permutation matrix is formed by taking the last coordinate of each $\mathcal{C}_i$, i.e., $(i - 1)m + a_s$'s as the diagonal.

References

- [ALWZ21] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795–815, 2021.
- [BCW21] Tolson Bell, Suchakree Chueluecha, and Lutz Warnke. Note on sunflowers. *Discrete Mathematics*, 344(11):112340, 2021.
- [BG25] Joshua Brakensiek and Venkatesan Guruswami. Redundancy is all you need. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 1614–1625, New York, NY, USA, 2025. Association for Computing Machinery.
- [BGP26] Joshua Brakensiek, Venkatesan Guruswami, and Aaron Putterman. Multiplicative error set system sparsification: A simpler proof via chain length contraction, 2026.
- [CKR22] Bruno Pasqualotto Cavalar, Mrinal Kumar, and Benjamin Rossman. Monotone circuit lower bounds from robust sunflowers. *Algorithmica*, 84(12):3655–3685, 2022.
- [CKST19] Charles Carlson, Alexandra Kolla, Nikhil Srivastava, and Luca Trevisan. Optimal lower bounds for sketching graph cuts. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2565–2569. SIAM, 2019.
- [ER60] P. Erdős and R. Rado. Intersection theorems for systems of sets. *Journal of the London Mathematical Society*, 35(1):85–90, 1960.
- [Fra82] P. Frankl. An extremal problem for two families of sets. *European Journal of Combinatorics*, 3(2):125–127, 1982.
- [Fra95] Peter Frankl. Extremal set systems. In *Handbook of Combinatorics*, volume 2, pages 1293–1329. Elsevier, 1995.
- [Fuk18] Junichiro Fukuyama. Improved bound on sets including no sunflower with three petals. arXiv preprint, 2018.
- [Gil22] Justin Gilmer. A constant lower bound for the union-closed sets conjecture, 2022.
- [KK07] Jeff Kahn and Gil Kalai. Thresholds and expectation thresholds. *Combinatorics, Probability and Computing*, 16(3):495–502, 5 2007.
- [Kos97] A. V. Kostochka. A bound of the cardinality of families not containing Δ -systems. In Ronald L. Graham and Jaroslav Nešetřil, editors, *The Mathematics of Paul Erdős II*, volume 14 of *Algorithms and Combinatorics*, pages 229–235. Springer, Berlin, Heidelberg, 1997.
- [KPS24] Sanjeev Khanna, Aaron (Louie) Putterman, and Madhu Sudan. Code sparsification and its applications. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms (SODA 2024)*, Alexandria, VA, USA, January 7–10, 2024, pages 5145–5168. SIAM, 2024.
- [KPS25] Sanjeev Khanna, Aaron Putterman, and Madhu Sudan. Efficient algorithms and new characterizations for csp sparsification. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 407–416, New York, NY, USA, 2025. Association for Computing Machinery.
- [LMM⁺22] Shachar Lovett, Raghu Meka, Ian Mertz, Toniann Pitassi, and Jiapeng Zhang. Lifting with Sunflowers. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, volume 215 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 104:1–104:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.
- [Neu28] John Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928.

- [PP24] Jinyoung Park and Huy Tuan Pham. A proof of the Kahn–Kalai conjecture. *Journal of the American Mathematical Society*, 37:235–243, 2024.
- [Rao20] Anup Rao. Coding for sunflowers. *Discrete Analysis*, 2020. Originally arXiv:1909.04774.
- [Ros14] Benjamin Rossman. The monotone complexity of k -clique on random graphs. *SIAM Journal on Computing*, 2014.
- [Sau72] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, July 1972.
- [Saw23] Will Sawin. An improved lower bound for the union-closed set conjecture, 2023.
- [She72] Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1):247–261, April 1972.

A Proof of Frankl’s theorem

In this section, we prove Theorem 3.2, restated below for convenience.

Theorem A.1 (Theorem 3.2 restated, [Fra82]). *Let $\mathcal{A} = \{A_1, \dots, A_m\}$ be a family of r -sets and $\mathcal{B} = \{B_1, \dots, B_m\}$ be a family of s -sets such that (1) $A_i \cap B_i = \emptyset$ for $i = 1, 2, \dots, m$ and (2) $A_i \cap B_j \neq \emptyset$ for $1 \leq i < j \leq m$. Then*

$$m \leq \binom{r+s}{s}.$$

The main idea behind the proof of Theorem 3.2 is as follows: we associate to each set A_i a vector u_i in some vector space W and show that $u_1, \dots, u_m$ are linearly independent. Given this, it suffices to choose W such that $\dim W = \binom{r+s}{s}$. In [Fra82], Frankl chose $W = \text{Sym}^r(V)$ for some vector space V of dimension $s+1$ where $\text{Sym}^r(V)$ denotes the r -th symmetric tensor power of V .

Definition A.2 (Symmetric tensor product). *Let $v_1, \dots, v_d \in V \subseteq \mathbb{R}^n$. The d -th symmetric tensor product of $v_1, \dots, v_d$ denoted $v_1 \vee v_2 \vee \dots \vee v_d$ is defined as*

$$v_1 \vee v_2 \vee \dots \vee v_d := \frac{1}{d!} \sum_{\sigma \in S_d} v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \dots \otimes v_{\sigma(d)}.$$

In other words, the symmetric tensor product is the average of tensor products over all permutations.

Definition A.3 (Symmetric tensor power). *Let $V \subseteq \mathbb{R}^n$. The d -th tensor power of V denoted $\text{Sym}^d(V)$ is defined as:*

$$\text{Sym}^d(V) := \text{span}\{v_1 \vee v_2 \vee \dots \vee v_d : v_i \in V, \forall i \in [d]\}.$$

Fact A.4 (Dimension of symmetric tensor power). *$\text{Sym}^d(V)$ is a vector space and*

$$\dim(\text{Sym}^d(V)) = \binom{\dim(V) + d - 1}{d}.$$

We say T is a w -uniform set if $|T| = w$. We are now ready to prove Theorem 3.2.

Proof. (of Theorem A.1, [Fra82]) Adding distinct elements and enlarging the universe if necessary, we may assume $\mathcal{A}$ is a and $\mathcal{B}$ are families of r -uniform and s -uniform sets respectively. Let $X := (\bigcup_i A_i) \cup (\bigcup_j B_j)$ denote the universe. Choose a vector space V of dimension $s+1$. Assign to each $x \in X$ a vector $v_x \in V$ such that any $s+1$ of these vectors are linearly independent⁸. For $A_i = \{a_{i,1}, \dots, a_{i,r}\}$, define u_i to be the associated symmetric r -th tensor product:

$$u_i := v_{a_{i,1}} \vee v_{a_{i,2}} \vee \dots \vee v_{a_{i,r}}.$$

⁸For instance, we can choose distinct real numbers t_x and set $v_x = (1, t_x, \dots, t_x^s) \in \mathbb{R}^{s+1}$. Then any $s+1$ vectors forms a Vandermonde matrix and are thus linearly independent.

Notice that for each j , the vectors $\{v_b : b \in B_j\}$ span an s -dimensional subspace of V as $|B_j| = s$. Since $\dim(V) = s + 1$, there exists a nonzero linear functional $f_j : V \rightarrow \mathbb{R}$ such that $f_j(v_b) = 0$ for all $v_b \in B_j$. We can extend f_j to obtain a functional $F_j : \text{Sym}^r(V) \rightarrow \mathbb{R}$ defined as

$$F_j(v_{x_1} \vee v_{x_2} \vee \cdots \vee v_{x_r}) = f_j(v_{x_1})f_j(v_{x_2}) \cdots f_j(v_{x_r}).$$

In particular, this implies $F_j(u_i) = \prod_{a \in A_i} f_j(v_a)$. Now, if $A_i \cap B_j \neq \emptyset$, then there exists some $a \in A_i \cap B_j$ which implies $f_j(v_a) = 0$ and hence $F_j(u_i) = 0$. On the other hand, if $A_i \cap B_j = \emptyset$ then for all $a \in A_i$, we have $|B_j \cup \{a\}| = s + 1$. By the construction of V , we know $\{v_b : b \in B_j\} \cup \{a\}$ is linearly independent for all $a \in A_i$. This gives $f_j(v_a) \neq 0$ for all $a \in A_i$ since f_j vanishes only on $\text{span}(\{v_b : v \in B_j\})$. Hence, in this case we have $F_j(u_i) \neq 0$.

Together with the assumptions on the set families $\mathcal{A}$ and $\mathcal{B}$, we thus know (1) $F_i(u_i) \neq 0$ for all $i = 1, 2, \dots, m$ and (2) $F_j(u_i) = 0$ for all $1 \leq i < j \leq m$. Using this, we can show $u_1, \dots, u_m$ are linearly independent.

To see this, suppose for contradiction $u_1, \dots, u_m$ are linearly dependent. Then there exists $c_1, \dots, c_m$ not all zeros such that

$$c_1 u_1 + \cdots + c_m u_m = 0.$$

Let $k \in [m]$ denote the largest index such that $c_k \neq 0$. Then we know $c_1 u_1 + \cdots + c_k u_k = 0$. Applying F_k to both sides, we have by linearity that

$$c_1 F_k(u_1) + \cdots + c_k F_k(u_k) = 0.$$

Since $F_k(u_i) = 0$ for all $i < k$, the above equality is equivalent to $c_k F_k(u_k) = 0$ which implies $c_k = 0$ as $F_k(u_k) \neq 0$. This is a contradiction.

Finally, since $u_1, \dots, u_m$ are linearly independent, we know

$$m \leq \dim(\text{Sym}^r(V)) \leq \binom{\dim(V) + r - 1}{r} = \binom{r + s}{s}$$

as desired where the last equality follows from Fact A.4. □