Electronic Colloquium on Computational Complexity, Report No. 69 (2026)

Moonflowers and efficient code sparsification

Shachar Lovett* UC San Diego

Raghu Meka† UCLA

Yimeng Wang‡ UCLA

Abstract

We introduce *moonflowers*, a weaker analogue of sunflowers. A family of sets $S_1, \ldots, S_k$ 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 near-optimal bounds.

As an application, we revisit the code sparsification problem studied by Brakensiek and Guruswami (STOC 2025) and improve the bounds to near optimal. Concretely, we improve the dependence on the block length from poly-logarithmic to logarithmic, and show that such a dependence 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 $S_1, \ldots, S_k$ is called a *k-sunflower* if their pairwise intersections are all the same; in other words, if if $S_i \cap S_j = S_1 \cap \cdots \cap S_k$ 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.

*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

ISSN 1433-8092

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.*

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 a 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, they were 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 $S_1, \ldots, S_k$ 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 $S_1 \setminus I, \ldots, S_k \setminus I$ 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 set[1]. 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-sunflower[2]. 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?*

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

[2] To 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.

We prove the following theorem.

Theorem 1.4 (Extremal bounds on moonflowers)**.** *There exists an absolute constant C such that the following holds. Let $k, w \geq 1$. Let $\mathcal{F}$ be a k-moonflower-free family of w-sets. Then,*

$$|\mathcal{F}| \leq \begin{cases} \left(C \cdot \frac{k}{w}\right)^w & \text{if } w \leq k, \\ \left(C \cdot \frac{w}{k}\right)^k & \text{if } w \geq k. \end{cases}$$

In particular, when $w = \Theta(k)$, the above bound simplifies to $|\mathcal{F}| \leq \exp(O(w))$. We complement this upper bound with an almost-matching lower bound via an explicit construction: take all subsets of size w from a universe of size $k + w - 2$. We obtain the following lemma.

Lemma 1.5 (Moonflower lower bound)**.** *There exists a family of w-sets $\mathcal{F}$ such that $|\mathcal{F}| = \binom{k+w-2}{w}$ and $\mathcal{F}$ is k-moonflower-free.*

Applying the standard bound $\binom{n}{m} \geq (n/m)^m$, we find the upper bound in Theorem 1.4 to be tight up to a constant factor in the base.

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 $\operatorname{supp}(\mathcal{F})$. 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)**.** *There exists an absolute constant C such that the following holds. Let $G = (L, R, E)$ be a w-left-regular 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*

$$|L| \leq \begin{cases} \left(C \cdot \frac{k}{w}\right)^w & \text{if } w \leq k, \\ \left(C \cdot \frac{w}{k}\right)^k & \text{if } w \geq k. \end{cases}$$

1.2 Code sparsification

Another major application of our moonflower bounds 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 $\alpha : T \to \mathbb{R}_{\geq 0}$. It induces the estimator

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

We say (T, α) *ε-sparsifies* $\mathcal{C}$ if $\widehat{\mathrm{wt}}_{T,\alpha}(x) \in (1 \pm \varepsilon)\mathrm{wt}(x)$ for all $x \in \mathcal{C}$ where $\mathrm{wt}(x) = \sum_{i=1}^n x_i$. 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 $\mathsf{poly}(n, 1/\varepsilon)$ time, one can compute an ε-sparsifer 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 $|T| \le O(\mathrm{NRD}(\mathcal{C})(\log n)^6/\varepsilon^2)$ where $\mathrm{NRD}(\mathcal{C})$ denotes the *non-redundancy* of $\mathcal{C}$ defined as follows.

Definition 1.7 ([BG25] Non-redundancy). *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 $\mathrm{NRD}(\mathcal{C})$, to be the size of the largest non-redundant set that is non-redundant for $\mathcal{C}$.*

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

Given a set family $\mathcal{F}$, let $\mathrm{MF}(\mathcal{F})$ denote the size of the largest moonflower contained in $\mathcal{F}$. Denote $\mathcal{F} := \{\mathrm{supp}(c) : c \in \mathcal{C}\} \subseteq 2^{[n]}$. Then we have $\mathrm{MF}(\mathcal{F}_C) = \mathrm{NRD}(C)$. 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 statement, see Lemma 2.12. In particular, this allows us to apply Theorem 1.4 to bound the size of $\mathcal{C}_{\le d} := \{c \in \mathcal{C} : \mathrm{wt}(c) \le d\}$ for any $d \in [n]$ as moonflower-freeness is preserved under projections. Using this and a refined analysis of the techniques introduced in [BG25], we are able to obtain the following result on code sparsification.

Theorem 1.8 (Improved code sparsification). *Let $\mathcal{C} \subseteq \{0,1\}^n$ with $\mathrm{NRD}(\mathcal{C}) = k$. Then for every $\varepsilon \in (0, 1/4)$ there exists a weighted coordinate set (T, α) that ε-sparsifies $\mathcal{C}$ and satisfies*

$$|T| \;\le\; \frac{k \log n}{\varepsilon^2} \cdot \mathsf{poly}\big(\log(k/\varepsilon), \log\log n\big).$$

Comparing with the result in [BG25], we bring the dependence on n to be optimal up to $\mathsf{poly}(\log\log n)$ factors instead of being off by $\mathsf{poly}(\log n)$ factors. As the following lemma shows, the $\log n$ dependence is necessary for ε-sparsifiers in general.

Lemma 1.9 (Lower bound on code sparsification). *Let $k \ge 1$ and $\varepsilon \in (0, 1)$. Then, for all large enough n, there exists an explicit $\mathcal{C} \subseteq \{0,1\}^n$ with $\mathrm{NRD}(\mathcal{C}) = k$ 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 1.8 up to polynomial factors in $1/\varepsilon$. We prove this in Section 4.3. It is an interesting question to see if the dependence on ε in the lower bound can be improved to $1/\varepsilon^2$, or if the upper bound dependency on ε can be improved.

1.3 Proof overview

In this subsection, we present the overview of our proofs. We first give a high-level idea of the proof of Theorem 1.4. While the overall structure of the proof is similar to that in [BG25], we simplify some of the proofs and make several important quantitative improvements to statements in [BG25]. Then via a carefully carried-out *iterative puncturing* argument, we obtain the desired bounds in Theorem 1.4. Next, we give an overview of the proof of Theorem 1.8. To obtain the optimal dependence on n, we crucially use the optimal moonflower bound to control the error probability of any codeword from random sampling. Finally, through an *iterative sampling* process and a refined analysis of [BG25], we are able to prove the improved sparsification result Theorem 1.8.

Moonflowers. For the purpose of the proof overview, we will focus on the case where $w \leq k$. The $w \geq k$ case follows the same idea but is slightly more technical. Let $\mathcal{F}$ be a family of w-sets over a universe U that does not contain any k-moonflower. We first claim that the support $\mathrm{supp}(\mathcal{F}) := \cup_{S \in \mathcal{F}} S$ cannot be too large. Concretely, $|\mathrm{supp}(\mathcal{F})| \leq (k-1)w$. To see this, consider the inclusion-minimal subfamily $\mathcal{F}' \subseteq \mathcal{F}$ subject to $\cup_{S \in \mathcal{F}'} S = U$. By minimality, $\mathcal{F}'$ is a moonflower. Applying the assumption that $\mathcal{F}$ is k-moonflower-free, we get $|U| \leq (k-1)w$ (for more details, see Lemma 2.9).

Using this, we are able to get an upper bound on $|\mathcal{F}|$: $|\mathcal{F}| \leq \binom{|U|}{\leq w} \leq (e|U|/w)^w = (ck)^w$ for some constant c. However, because of the dependence on k in the base, this upper bound is far from being satisfactory. The key deficiency of the argument above comes from the upper bound $|U| \leq (k-1)w$. If we can somehow *reduce the size of the universe while keeping most of the sets in* $\mathcal{F}$, then we can use the same argument to achieve a much better bound.

To make this precise, suppose we are able to identify a popular subset $I \subseteq U$ such that (i) $|I| \leq t$ and (ii) $|\{S \in \mathcal{F} : S \subseteq I\}| \geq (1-\eta)|\mathcal{F}|$. Then $|\mathcal{F}| \leq (1-\eta)^{-1}|\{S \in \mathcal{F} : S \subseteq I\}| \leq (1-\eta)^{-1}\binom{|I|}{\leq w}$. In the case where $1-\eta = \Omega(1)$ and $|I| = O(k)$, this yields the desired bound $|\mathcal{F}| \leq (Ck/w)^w$. So the question now becomes: *how can we obtain such a subset?*

To answer this question, 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 $S \in \mathcal{F}$, $\mathbf{Pr}_{i \sim Q}[i \in S] \geq p$. 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 $\mathcal{F}_J \backslash \mathcal{S}_J$ is p-covered. We say $\mathcal{F}$ is *(p, M)-almost-covered* if it satisfies this condition. With this in hand, we are able to prove our one-step universe reduction theorem.

Theorem 1.10 (One-step universe reduction of almost-covered-families)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w-sets. Fix $p \in (0,1), M \geq 1$ and $\delta \in (0, 1/2)$. If $\mathcal{F}$ is (p, M)-almost-covered, then there exists an $I \subseteq [n]$ with $|I| \leq t$ such that*

$$\left|\left\{S \in \mathcal{F} : S \subseteq I\right\}\right| \geq (1-\delta)|\mathcal{F}| - tM,$$

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

Notice Theorem 1.10 allows us to identify the important subset I as desired. More importantly, once we identify such an I and restrict out family to be $\mathcal{F}' := \{S \in \mathcal{F} : S \subseteq I\}$, we can iteratively apply Theorem 1.10. With appropriately chosen parameters in each iteration, we eventually prove Theorem 1.4 via *iterative puncturing*.

The last piece of the puzzle is how we can bound the parameter M. Indeed Theorem 1.10 is only interesting if M is small. To establish this, we crucially rely on Gilmer's entropy method [Gil22, Saw23] as used in [BG25]. Together with [Sau72, Saw23] and the k-moonflower-freeness of $\mathcal{F}$, we show the following lemma.

Lemma 1.11 (Moonflower-free families are almost covered)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k-moonflower-free family of w-sets. Assume $n = |\mathrm{supp}(\mathcal{F})|$. Then $\mathcal{F}$ is $(p, 2^h)$-almost-covered where*

$$h = h(n, k, p) := B \cdot k \log\left(\frac{n}{k}\right) \cdot p \log(1/p),$$

for some absolute constant B.

Finally, as mentioned earlier, since $\mathcal{F}$ is k-moonflower-free, we have $n \leq (k-1)w$. As a result, h depends only on k, w, p. We present the full proof in Section 3.

Code sparsification. Now we 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 sparsifer 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 $\mathrm{wt}(c) = 1$. Then the only way for c's weight to be preserved is to include the sole coordinate in $\mathrm{supp}(c)$ 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 sparsifer. 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 $\bigcup_{c \in \mathcal{C}_{\leq w_{\min}}} \mathrm{supp}(c)$ into the sparsifer T and assign each such coordinate a weight of 1. Then we add each coordinate $i \in [n] \backslash (\bigcup_{c \in \mathcal{C}_{\leq w_{\min}}} \mathrm{supp}(c))$ 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 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 1.10 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 $\mathcal{F}_{\mathcal{C}} := \{\mathrm{supp}(c) : c \in \mathcal{C}\}$. We make the observation that $\mathrm{NRD}(\mathcal{C}) = \mathrm{MF}(\mathcal{F}_{\mathcal{C}})$ where $\mathrm{MF}(\mathcal{G})$ denotes the size of the largest moonflower contained in set family $\mathcal{G}$. For a proof of this, see Lemma 2.12. We divide the weights into three regimes: low-weight ($|c| \leq w_{\min}$); medium-weight ($w_{\min} \leq |c| \leq w_*$) and high-weight ($|c| \geq w_*$) regimes. For the low-weight regime, we again include the support of all low-weight codewords into the sparsfier. 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 $w \leq w_*$ and include the other coordinates independently with probability 1/2. The reason for this is the following: the trace bound we obtain from Theorem 1.10 deteriorates as the weight increases. At that point, the bound $|\mathcal{F}_{\le w}| \le (Ck/w)^w$ we obtain from the optimal moonflower bound in Theorem 1.4 outperforms the bound in Theorem 1.10.

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 the our improved universe-reduction bound Theorem 1.10 and the moonflower bound Theorem 1.4 with appropriately chosen parameters, we are able to obtain a sparsifier with the guarantees in Theorem 1.8. For a more thorough discussion of our improved sparsification strategy, see Section 4.2.

1.4 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 bounds for the size of families of w-sets without moonflowers. In Section 4, we use this tight bound to prove our code sparsification theorem.

1.5 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? The main difficulty seems to be on how to algorithmically remove M sets from a set family $\mathcal{F}$ so that the remaining set family is p-covered.

(2) **Better sparsification lower bound.** Our current code sparsification lower bound is near optimal in terms of its dependence on $\mathrm{NRD}(\mathcal{C})$ 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.

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 $[n] := \{1, 2, ..., n\}$. 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 $\bar{I} := [n]\backslash I$. We use $2^{[n]}$ to denote the family of subsets of $[n]$.

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

Definition 2.1 (Moonflower). *A family of sets $S_1, ..., S_k \subseteq [n]$ is a k-moonflower if there exists $I \subseteq [n]$ such that the sets $S_i\backslash I$ 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 $S_1, ..., S_k$ 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 $\mathrm{MF}(\mathcal{F})$ to be the largest k such that $\mathcal{F}$ contains a k-moonflower.

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

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

Definition 2.3 (Projection / trace). *For $\mathcal{F} \subseteq 2^{[n]}$ and $J \subseteq [n]$, define*

$$\mathcal{F}_J := \{S \cap J : S \in \mathcal{F}\} \subseteq 2^J.$$

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

Definition 2.5 (Coveredness). *We say a set $S \in \mathcal{F}$ is* covered *by $I \subseteq [n]$ if $S \subseteq I$.*

Lemma 2.6 (Size of restricted family). *Let $\mathcal{F} \subseteq 2^{[n]}$ and $I \subseteq [n]$, then $|\mathcal{F}_I| \geq |\mathcal{F}|/2^{|\bar{I}|}$.*

Proof. Since $\mathcal{F} \subseteq \mathcal{F}_I \times \mathcal{F}_{\bar{I}}$, we know $|\mathcal{F}| \leq |\mathcal{F}_I| \times |\mathcal{F}_{\bar{I}}| \leq |\mathcal{F}_I| \cdot 2^{|\bar{I}|}$. Rearranging gives the desired result. □

Definition 2.7 (Union-closure). *The* union-closure *of $\mathcal{F}$ is*

$$\mathcal{U}(\mathcal{F}) := \Big\{ \bigcup_{t=1}^{r} S_t : r \geq 0, S_t \in \mathcal{F} \Big\}.$$

Definition 2.8 (Shattering and VC-dimension). *Let $\mathcal{F} \subseteq 2^{[n]}$. A set $S \subseteq [n]$ is* shattered *by $\mathcal{F}$ if*

$$\{T \cap S : T \in \mathcal{F}\} = 2^S.$$

The VC dimension *of $\mathcal{F}$ is*

$$VC(\mathcal{F}) := \sup\{|S| :\ S \subseteq [n] \text{ and } S \text{ is shattered by } \mathcal{F}\}.$$

Lemma 2.9 (Support bound from moonflower-freeness). *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k-moonflower-free family of w-sets. Then*

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

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

Proof. If $\mathcal{F} = \emptyset$ then $\mathrm{supp}(\mathcal{F}) = \emptyset$ and we are done. Let $U := \mathrm{supp}(\mathcal{F})$. 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

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

as required. □

Lemma 2.10 (Lemma 1.5, restated). *There exists a family of w-sets $\mathcal{F}$ such that $|\mathcal{F}| = \binom{k+w-2}{w}$ and $\mathcal{F}$ is k-moonflower-free. In particular, this implies $\mathrm{MF}(k,w) \geq \binom{k+w-2}{w}$.*

Proof. Consider the following $\mathcal{F}$:

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

In words, $\mathcal{F}$ is the family of all w-sets from a universe of size $k+w-2$. Clearly, $|\mathcal{F}| = \binom{k+w-2}{w}$. It remains to show $\mathcal{F}$ is indeed k-moonflower free.

To see this, suppose $\mathcal{F}$ contains a k-moonflower with petals $S_1, ..., S_k$ and core $I \subseteq [k+w-2]$. Let $\ell = \min_{i=1,...,k} |S_i \backslash I| \geq 1$. Then, by the definition of k-moonflower, we have

$$\Big|\bigcup_{i=1}^{k} S_i\Big| = |I| + \sum_{i=1}^{k} |S_i \backslash I| \geq |I| + k\ell \geq (w-\ell) + k\ell \geq k+w-1$$

where the last inequality holds since $S_i \backslash I$ is non-empty for each i. Notice the inequality is tight if we take $S_i = \{1, 2, ..., w-1, w-1+i\}$ for $i = 1, 2, ..., k$. We have thus reached a contradiction since the universe is of size $k+w-2 < k+w-1$. □

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 $\mathrm{supp}(c) \subseteq [n]$ to be the set of nonzero coordinates. Define $\mathrm{supp}(\mathcal{C}) := \bigcup_{c\in C} \mathrm{supp}(c)$. For any codeword $c \in \mathcal{C}$, we define its *Hamming weight* as $|c| := c_1 + ... + c_n$, i.e., the number of nonzero coordinates. For any $d \in [n]$, let $\mathcal{C}_{\leq d}$ be the set of codewords of $\mathcal{C}$ 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 2.11 ([BG25] Non-redundancy)**.** *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 $\mathrm{NRD}(\mathcal{C})$, 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 $\mathrm{NRD}(\mathcal{C})$ is the dimension of the largest permutation submatrix contained in $\mathcal{C}$. When $\mathcal{C}$ is a linear code, $\mathrm{NRD}(\mathcal{C})$ equals the dimension of $\mathcal{C}$. In addition, for any non-trivial $\mathcal{C}$, we have $1 \leq \mathrm{NRD}(\mathcal{C}) \leq n$. Let $\mathcal{F}_{\mathcal{C}} := \{\mathrm{supp}(c) : c \in \mathcal{C}\} \subseteq 2^{[n]}$. We have the following simple lemma establishing a connection between $\mathrm{NRD}(\mathcal{C})$ and $\mathrm{MF}(\mathcal{F}_{\mathcal{C}})$.

Lemma 2.12 (NRD equals MF of the support family)**.** *Let $\mathcal{C} \subseteq \{0,1\}^n$ be a binary code. Then*

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

Proof. We prove the two inequalities.
(1) $\mathrm{MF}(\mathcal{F}_{\mathcal{C}}) \geq \mathrm{NRD}(\mathcal{C})$**.** Let $I \subseteq [n]$ be non-redundant for $\mathcal{C}$ with $|I| = \mathrm{NRD}(\mathcal{C})$. 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, $\mathrm{supp}(c^{(i)}) \cap I = \{i\}$. Let $S_i := \mathrm{supp}(c^{(i)}) \in \mathcal{F}_{\mathcal{C}}$ 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 $\mathrm{MF}(\mathcal{F}_{\mathcal{C}}) \geq |I| = \mathrm{NRD}(\mathcal{C})$.
(2) $\mathrm{NRD}(\mathcal{C}) \geq \mathrm{MF}(\mathcal{F}_{\mathcal{C}})$**.** Let $k := \mathrm{MF}(\mathcal{F}_{\mathcal{C}})$, so $\mathcal{F}_{\mathcal{C}}$ contains a k-moonflower $S_1, \ldots, S_k$ 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 $I := \{i_1, \ldots, i_k\}$. For each t, let $c^{(t)} \in \mathcal{C}$ be a codeword with $\mathrm{supp}(c^{(t)}) = S_t$. We claim that $\mathrm{supp}(c^{(t)}) \cap I = \{i_t\}$. 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 $\mathrm{NRD}(\mathcal{C}) \geq |I| = k = \mathrm{MF}(\mathcal{F}_{\mathcal{C}})$.

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

2.3 Chernoff bound

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

Lemma 2.13 (Chernoff bound, convenient form)**.** *Let $t \geq 1$ and let $X \sim \mathrm{Bin}(t, 1/2)$. Then for every $\Delta > 0$,*

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

Proof. Standard Chernoff bound in additive form. □

3 Bounding the size of set families without a moonflower

In this section, we prove our main result on moonflowers, giving a tight upper bound to $\mathrm{MF}(k, w)$.

Theorem 3.1 (Theorem 1.4 restated)**.** *There exists an absolute constant C such that the following holds. Let $k, w \geq 1$. Let $\mathcal{F}$ be a k-moonflower-free family of w-sets. Then,*

$$|\mathcal{F}| \leq \begin{cases} \left(C \cdot \frac{k}{w}\right)^w & \text{if } w \leq k \\ \left(C \cdot \frac{w}{k}\right)^k & \text{if } w \geq k. \end{cases}$$

In particular, if $w = \Theta(k)$, then $|\mathcal{F}| \leq \exp(O(w))$.

Let's examine the tightness of the above bound: by Lemma 2.10, we know

$$\mathrm{MF}(k, w) \geq \binom{k+w-2}{w} \approx \binom{k+w}{w} \geq \max\left\{\left(\frac{k+w}{w}\right)^w, \left(\frac{k+w}{k}\right)^k\right\}$$
$$= \max\left\{\left(1+\frac{k}{w}\right)^w, \left(1+\frac{w}{k}\right)^k\right\}.$$

This implies we have obtained matching upper and lower bounds in all parameter regimes up to constant factors in the base.

We will first focus on the proof in the $w \leq k$ case. Then we slightly modify the argument to prove the $w \geq k$ case. The proof follows the high-level approach of [BG25] where several aspects of the proof are streamlined and optimized. Before we start with the proof, we need to make two definitions that will play important roles in the proof of Theorem 3.1.

Definition 3.2 (p-smooth distribution)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$. A probability distribution D supported on $\mathcal{F}$ is* p-smooth *if for every coordinate $i \in [n]$,*

$$\mathop{\mathbf{Pr}}_{S \sim D}[i \in S] \leq p.$$

Definition 3.3 (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 $S \in \mathcal{G}$,*

$$\mathop{\mathbf{Pr}}_{i \sim Q}[i \in S] \geq p.$$

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.

On a very high-level, the main idea behind the proof Theorem 3.1 is to reduce the universe size to be roughly $O(k)$ if $k \geq w$ or $O(w)$ if $w \geq k$. For this outline, let us focus on the case $k \gg w$. Let $\mathcal{F}$ be a k-moonflower-free family of w-sets. We will show that there exists a set I of size roughly $O(k)$ such that $|\mathcal{F}_I| \geq 2^{-O(w)}|\mathcal{F}|$. The theorem now follows from this immediately by counting the number of w-sets.

We next describe how to find the set I so that $|I|$ is small but $|\mathcal{F}_I|$ is large. We do this iteratively, by starting with the easy bound that $n \leq kw$. This can be broken into four somewhat modular steps:

1. Following the arguments of [BG25], we show that any p-smooth distribution D supported on a k-moonflower free family $\mathcal{F}$ has entropy at most $O(k \cdot \log(n/k) \cdot p \log(1/p))$. This step critically uses the entropy-amplification property of taking unions proved by [Gil22, Saw23]. See Theorem 3.4.

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 Theorem 3.9.

3. Finally, suppose 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, there exists a small set $I \subseteq [n]$, $|I| = O(w/p)$ such that $|\mathcal{F}_{\overline{I}}| \leq |I| \cdot 2^H$. In other words, this also implies that $|\mathcal{F}_I|$ is quite large compared to $|\mathcal{F}|$. See Theorem 3.12 for the precise version.

4. Once we have the above *single-step* universe reduction argument, we iterate it carefully to get our final bounds.

The first two steps are implicit in the arguments of [BG25]; we abstract this way to streamline the argument and get the quantitative improvements needed for step three. Iterating the argument is critical for getting the final tight bounds.

3.1 Entropy bound for smooth distributions

In this subsection, we prove our improved smoothness bound. For $p \in (0,1)$, define $\phi(p) := p\log(1/p)$.

Theorem 3.4 (Improved smoothness lemma)**.** *There exists an absolute constant $B \geq 1$ such that the following holds. Let $\mathcal{F} \subseteq 2^{[n]}$ be k-moonflower-free. Then for every p-smooth distribution D supported on $\mathcal{F}$,*

$$H(D) \leq B \cdot k\log(n/k) \cdot \phi(p).$$

Towards a proof of Theorem 3.4, we first show that for any set family $\mathcal{F}$, we can upper bound the VC-dimension of its union closure using $\mathrm{MF}(\mathcal{F})$.

Lemma 3.5 (Moonflower $\implies$ small VC of union-closure, [BG25, Proposition 4.3], restated)**.** *Let $\mathcal{F}$ be a family of sets. If $\mathcal{F}$ is k-moonflower-free, then* $\mathrm{VC}(\mathcal{U}(\mathcal{F})) \leq k-1$.

Proof. Suppose for contradiction that $\mathrm{VC}(\mathcal{U}(\mathcal{F})) \geq k$. Then there exists $J \subseteq [n]$ with $|J| = k$ that is shattered by $\mathcal{U}(\mathcal{F})$; i.e., for every $T \subseteq J$, there exists $U_T \in \mathcal{U}(\mathcal{F})$ with $U_T \cap J = T$.

In particular, for each $j \in J$ there exists $U_{\{j\}} \in \mathcal{U}(\mathcal{F})$ satisfying $U_{\{j\}} \cap J = \{j\}$. Since $U_{\{j\}}$ is in the union-closure of $\mathcal{F}$, we can write $U_{\{j\}} = \bigcup_{t=1}^{r_j} S_{j,t}$ for some $S_{j,t} \in \mathcal{F}$ and $r_j > 0$. Because $j \in U_{\{j\}}$, at least one constituent set, call it S_j, contains j. Moreover, as $U_{\{j\}} \cap (J\setminus\{j\}) = \emptyset$, every constituent $S_{j,t}$ must satisfy $S_{j,t} \cap (J\setminus\{j\}) = \emptyset$. In particular, we have $S_j \cap J = \{j\}$.

Let $I := [n]\setminus J$. Then for each $j \in J$,

$$S_j\setminus I = S_j \cap J = \{j\}.$$

This implies that $\{S_j : j \in J\}$ forms a k-moonflower in $\mathcal{F}$ with core I contradicting the assumption. As a result, $\mathrm{VC}(\mathcal{U}(\mathcal{F})) \leq k-1$. ☐

The next step is to upper bound the size of the set family using its VC-dimension.

Lemma 3.6 (Upper bound from Sauer-Shelah [Sau72, She72])**.** *Let $\mathcal{H} \subseteq 2^{[n]}$ satisfying* $\mathrm{VC}(\mathcal{H}) \leq d$. *Then*

$$|\mathcal{H}| \leq \sum_{i=1}^{d} \binom{n}{i}.$$

Consequently, for all $1 \leq d \leq n$,

$$\log|\mathcal{H}| \leq O(d\log(n/d)).$$

Proof. The Sauer-Shelah lemma [Sau72, She72] gives $|\mathcal{H}| \le \sum_{i=1}^{d} \binom{n}{i}$.

If $d \le n/2$, then $\binom{n}{i}$ is non-decreasing for $0 \le i \le d$. Hence,

$$\sum_{i=0}^{d} \binom{n}{i} \le (d+1)\binom{n}{d} \le (d+1)\Big(\frac{en}{d}\Big)^{d}.$$

Taking logarithms on both sides gives

$$\log|\mathcal{H}| \le d \log\left(\frac{en}{d}\right) + \log(d+1).$$

If $n/2 < d \le n$, then since $|\mathcal{H}| \le 2^n$, we have $\log|\mathcal{H}| \le n < 2d$. On the other hand, $\log(en/d) \ge \log e > 1$ as $en/d \ge e$ when $d \le n$. Consequently, $\log|\mathcal{H}| < 2d\log(en/d)$. Combining the two cases yields the desired result. □

It remains to show how to bound the entropy of a p-smooth distribution. To this end, we use Gilmer/Sawin's entropy bound [Gil22, Saw23] as a black box. Below we state the precise statement we need. For a proof of the statement, see [BG25, Corollary 4.13].

Lemma 3.7 (Gilmer/Sawin entropy amplification, [BG25, Corollary 4.13])**.** *For $p \in (0,1)$, define $\phi(p) := p\log(1/p)$. Let $\mathcal{F} \subseteq 2^{[n]}$ be any finite family, and let D be a p-smooth distribution supported on $\mathcal{F}$. Then*

$$H(D) \le C_{GS} \cdot \phi(p) \cdot \log|\mathcal{U}(\mathcal{F})|,$$

where $H(\cdot)$ denotes the Shannon entropy, and $C_{GS} > 0$ is an absolute constant.

The following lemma allows us to upper bound $\phi(p)$ via a convenient choice of p.

Lemma 3.8 (A convenient choice of p)**.** *Let $a \in (0, 1/4]$ and define*

$$p \ := \ \frac{a}{4\log(4/a)}.$$

Then $p \in (0, 1/2]$ and

$$\phi(p) \le a.$$

Proof. Since $a \le 1/4$, we have $\log(4/a) \ge 2$, hence $p \le a/8 \le 1/32$. Also,

$$\log\Big(\frac{1}{p}\Big) = \log\Big(\frac{4\log(4/a)}{a}\Big) = \log\Big(\frac{4}{a}\Big) + \log\big(\log(4/a)\big).$$

For $a \le 1/4$, $\log(\log(4/a)) \le \log(4/a)$, so $\log(1/p) \le 2\log(4/a)$. Therefore

$$p\log(1/p) \ \le \ \frac{a}{4\log(4/a)} \cdot 2\log(4/a) \ = \ \frac{a}{2} \ \le \ a.$$

□

We now state and prove our improved smoothness lemma.

Proof of Theorem 3.4. By Lemma 3.5, we know $\mathrm{VC}(\mathcal{U}(\mathcal{F})) \le k-1$. Applying Lemma 3.6 with $d = k-1$ yields

$$\log|\mathcal{U}(\mathcal{F})| \le O\Big((k-1)\log\left(\frac{en}{k-1}\right)\Big) = O\Big(k\log(n/k)\Big).$$

Plugging this in Lemma 3.7, we find

$$H(D) \le B \cdot k\log(n/k) \cdot \phi(p)$$

where we absorb the constants into B. □

3.2 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 3.9 (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 $\mathcal{F} \backslash \mathcal{S}$ 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 3.10 (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

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

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 $\mathbf{Pr}_{T \sim D}[i \in T]$. The column player best-responds with $\max_i \mathbf{Pr}_{T \sim D}[i \in T]$. The row player then minimizes this quantity.

By von Neumann's minimax theorem [Neu28],

$$\max_{Q} \min_{T \in \mathcal{G}} \sum_{i \in T} Q(i) = \min_{D} \max_{i \in [n]} \mathop{\mathbf{Pr}}_{T \in D}[i \in T],$$

as desired. □

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 3.10 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 3.9. If $\mathcal{F}$ is already p-covered, take $S = \emptyset$ and we are done. Otherwise by Proposition 3.10 (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) $\mathcal{F}\backslash\mathcal{S}$ is p-covered and (2) $|\mathcal{S}| \leq 2^H$.

Step 1: $\mathcal{F}\backslash\mathcal{S}$ is p-covered. If $\mathcal{F}\backslash\mathcal{S} = \emptyset$, this is trivially true. Otherwise, suppose for contradiction that $\mathcal{F}\backslash\mathcal{S}$ is not p-covered. Then Proposition 3.10(2) yields a p-smooth distribution μ supported on $\mathcal{F}\backslash\mathcal{S}$. For $\epsilon \in (0, 1)$, define $\nu_\epsilon := (1-\epsilon)\nu^* + \epsilon\mu$. Since this is a convex combination of p-smooth distributions over $\mathcal{F}$, we have $\nu_\epsilon \in \mathcal{P}$. Now,

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

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

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

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

  for all $T \in \mathcal{F}\backslash\mathcal{S}$, contradicting the optimality of ν^*. Hence, $\mathcal{F}\backslash\mathcal{S}$ 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 $|S| \leq 1/\tau^*$. 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, $|S| \leq 2^H$.

□

3.3 One-step puncturing

We now describe what we call a *puncturing* step. While $\mathcal{F}$ is supported on $[n]$, we show that under certain conditions, we can identify a small universe $I \subseteq [n]$ such that most sets in $\mathcal{F}$ are contained in I, and so we can reduce the universe from $[n]$ to I while preserving most sets. We start by making the following definitions:

Definition 3.11 ((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 $\mathcal{F}_J\backslash\mathcal{S}_J$ 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 $\mathcal{F}$.

Theorem 3.12 (Theorem 1.10 restated)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w-sets. Fix $p \in (0, 1), M \geq 1$ and $\delta \in (0, 1/2)$. If $\mathcal{F}$ is (p, M)-almost-covered, then there exists an $I \subseteq [n]$ with $|I| \leq t$ such that*

$$\left|\left\{S \in \mathcal{F} : S \subseteq I\right\}\right| \geq (1-\delta)|\mathcal{F}| - tM,$$

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

In words, the theorem 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 3.11. 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 $I_0 := \emptyset, J_0 = [n], \mathcal{S}^{(0)}_{\text{removed}} = \emptyset$ and $\mathcal{F}_0 = \mathcal{F}$.
- For $j = 0, 1, ..., t_{\text{end}} - 1$:
  - Let $(\mathcal{S}_{J_j}, Q_{J_j})$ denote the corresponding witness for J_j.
  - Remove exceptions: $\mathcal{F}^-_j := \mathcal{F}_j \backslash \mathcal{S}_{J_j}$.
  - Sample $i_{j+1} \sim Q_{J_j}$ and add it: $I_{j+1} = I_j \cup \{i_{j+1}\}$.
  - Update $J_{j+1} := [n] \backslash I_{j+1}$ and $\mathcal{S}^{(j+1)}_{\text{removed}} = \mathcal{S}^{(j)}_{\text{removed}} \cup \mathcal{S}_{J_j}$.
  - Set $\mathcal{F}_{j+1} = (\mathcal{F}^-_j)_{J_{j+1}}$, and remove any empty sets or duplicate sets if some exist.

Notice for each j, $\mathcal{F}_j$ is precisely the family of sets in $\mathcal{F} \backslash \mathcal{S}^{(j)}_{\text{removed}}$ that are not covered by I_j. Now define the potential function[3]:

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

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 $|\mathcal{F}_j| \geq M$. Let's assume this is always the case before we finish (we will justify this assumption later). Now condition on the choice of i_k up to iteration 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 $\mathbf{Pr}_{i_{j+1} \sim Q_{J_j}}[i_{j+1} \in A] \geq p$. Moreover, we know

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

Therefore,

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

By linearity of expectation, we thus find

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

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

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

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

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 $I = \{i_1, \ldots, i_{t_{\text{end}}}\}$ 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

$$\left|\left\{S \in \mathcal{F} : S \not\subseteq I\right\}\right| \leq |\mathcal{F}_{t_{\text{end}}}| + |\mathcal{S}_{\text{removed}}^{t_{\text{end}}}| \leq \delta|F| + t_{\text{end}} M.$$

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. □

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 3.13 (Lemma 1.11, restated)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k-moonflower-free family of w-sets. Assume $n = |\text{supp}(\mathcal{F})|$. Then $\mathcal{F}$ is $(p, 2^h)$-almost-covered where*

$$h = h(n, k, p) := B \cdot k \log(n/k) \cdot p \log(1/p),$$

for some absolute constant $B \geq 1$.

Proof. Trimming unused elements in the universe if necessary, we can assume $n = |\text{supp}(\mathcal{F})|$. Since $\mathcal{F}$ is k-moonflower-free, by Lemma 2.4, each projection $\mathcal{F}_J$ is also k-moonflower-free. Moreover, for any $J \subseteq [n]$, we have $\log(|J|/k) \leq \log(n/k)$. By Theorem 3.4, every p-smooth distribution D supported on $\mathcal{F}_J$ must satisfy

$$H(D) \leq B \cdot k \log(n/k) \cdot p \log(1/p) = h,$$

where $B \geq 1$ is an absolute constant. Now by Theorem 3.9, for each J there exists an exceptional set $\mathcal{S}_J \subseteq \mathcal{F}_J$ with $|\mathcal{S}_J| \leq 2^h$ such that $\mathcal{F}_J \backslash \mathcal{S}_J$ is p-covered. In particular, this implies $\mathcal{F}$ is $(p, 2^h)$-almost-covered.

□

3.4 Iterated puncturing: $w \leq k$ case

In this section, we prove our main theorem (Theorem 3.1) in the case where $w \leq k$.

Theorem 3.14 (Iterated puncturing when $w \leq k$)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k-moonflower-free family of w-sets such that $w \leq k$. Then $|\mathcal{F}| \leq (Ck/w)^w$ for some absolute constant C.*

As mentioned earlier, our proof follows an iterated approach. Starting with a family of w-sets $\mathcal{F}$ supported on $[n]$, at each step, by Lemma 3.15, we either have the desired upper bound $(Ck/w)^w$ or identify a set of important elements $I \subseteq [n]$ of size $O(k \log(n/k) \log\log(n/k))$ such that $|\mathcal{F}| \geq (1 - 2^{-w})|\mathcal{F}|$. In the former case, since our procedure always terminates before losing at most a constant fraction of $\mathcal{F}$, we obtain our desired bound.

In the latter case, notice that $|I|/k = O(\log(n/k)\log\log(n/k))$ and $n/k \le w$ via Lemma 2.9. Hence, within $\log^* w$ iterations we will obtain a set of coordinates of size $O(k)$. In particular, since $\log^* w \ll 2^w$, we will only lose at most a constant fraction of $\mathcal{F}$ during this entire process. This in turn, enables us to obtain our final upper bound on $|\mathcal{F}|$.

To make the whole proof concrete, we first need the following one-step universe reduction lemma. Recall that for $p \in (0,1)$ we defined $\phi(p) = p\log(1/p)$.

Lemma 3.15 (One-step universe reduction when $w \le k$)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k-moonflower-free family of w-sets such that $w \le k$. Then there exists some absolute constant $C > 0$ such that either*

(i) $|\mathcal{F}| \le (Ck/w)^w$ or

(ii) There exists $I \subseteq [n]$ of size $|I| = t$ such that $|\{S \in \mathcal{F} : S \subseteq I\}| \ge (1 - 2^{-w})|\mathcal{F}|$ holds,

where $t := \min\{(k-1)w, Cw/p\}$. Here $p \in (0,1)$ is chosen such that $h(n,k,p) \le O(w\log(k/w))$. In particular, p is of order

$$p = \Theta\Big(\frac{w\log(k/w)}{k\log(n/k)\log(\frac{k\log(n/k)}{w\log(k/w)})}\Big). \tag{1}$$

Proof. By Lemma 2.9, we can truncate the universe to $U_0 = \mathrm{supp}(\mathcal{F})$ so that $n_0 := |U_0| \le (k-1)w$. By Lemma 3.13, we know $\mathcal{F}$ is $(p, 2^{h(n_0,k,p)})$-almost-covered where $h(n_0,k,p) := B \cdot k\log(n_0/k)\cdot\phi(p)$. Choose p such that

$$h(n_0,k,p) := B \cdot k\log(n_0/k)\cdot\phi(p) \le \frac{w\log(4k/w)}{32}. \tag{2}$$

We will justify why this is possible at the end of the proof. With this choice of p, we know $\mathcal{F}$ is (p, M)-almost-covered for $M \le (4k/w)^{w/32}$. Applying Theorem 3.12 with $\delta = 2^{-w-4}$, there exists an $I \subseteq [n]$ with $|I| \le t := \min\{(k-1)w, \lceil \frac{2}{p}(w\ln 2 + w + 4)\rceil\}$ such that

$$|\{S \in \mathcal{F} : S \subseteq I\}| \ge (1 - 2^{-w-4})|\mathcal{F}| - tM.$$

Now if $tM \le \delta|\mathcal{F}|$, then

$$|\{S \in \mathcal{F} : S \subseteq I\}| \ge (1 - 2\delta)|\mathcal{F}| \ge (1 - 2^{-w})|\mathcal{F}|$$

since $2\delta = 2^{-w-3} \le 2^{-w}$. Otherwise, we have $tM > \delta|\mathcal{F}|$. Using $M \le (4k/w)^{w/32}$ and rearranging, we obtain the desired result in case (i).

Finally we justify the choice of p. Let $p := a/(4\log(4/a))$ where a is defined as

$$a := \frac{w\log(4k/w)}{32Bk\log(n/k)} \in (0, 1/4].$$

To see why this quantity is bounded above by $1/4$, notice that when $w \le k$, we have $(w/k)\log(4k/w) \le \log 4$ via a routine calculation. Hence, $a \le \log 4/(32B\log(n/k)) < 1/4$. In particular, we have $p \in (0,1]$. By Lemma 3.8, $p\log(1/p) \le a$. This implies

$$h(n,k,p) := B \cdot k\log(n/k)\cdot\phi(p) \le B \cdot k\log(n/k)\cdot\frac{w\log(4k/w)}{32Bk\log(n/k)} = \frac{w\log(4k/w)}{32}.$$

$\square$

We are now ready to formally prove the main theorem of this section.

Proof of Theorem 3.14. It is easy to see the theorem holds for $w = 1$: any k distinct singleton sets form a k-moonflower. From now on, let's assume $2 \leq w \leq k$. By Lemma 2.9, we can truncate the universe to $U_0 = \text{supp}(\mathcal{F})$ so that $n_0 := |U_0| \leq (k-1)w$. Consider the following process:

Iterated puncturing: let $\mathcal{F}_0 = \mathcal{F}$ and $U_0 = \text{supp}(\mathcal{F})$. Let i_{end} be the stopping time which is to be determined by some stopping condition. For $i = 0, 1, 2, ..., i_{\text{end}}$:

- Let U_i denote the reduced universe at iteration i. Then $\mathcal{F}_i \subseteq 2^{U_i}$.
- Define $n_i := |U_i|$. Let $p_i \in (0, 1)$ be such that

  $$h(n_i, k, p_i) := B \cdot k \log(n_i/k) \cdot \phi(p_i) = w \log(4k/w)/32.$$

  as in (2).
- Assuming $\mathcal{F}_i$ is k-moonflower-free (we will prove this indeed is the case in Claim 3.16(1)), then by Lemma 3.13, we know $\mathcal{F}_i$ is $(p_i, 2^{h(n_i,k,p_i)})$-almost-covered. In particular, we have $2^{h(n_i,k,p_i)} \leq (4k/w)^{w/32}$ for all i. Applying Lemma 3.15 with $p = p_i$ and universe U_i, we either (i) obtain an upper bound on the family size: $|\mathcal{F}_i| \leq (Ck/w)^w$ or (ii) there exists a set $I_i \subseteq U_i$ with $|I_i| = t_i := \min\{(k-1)w, Cw/p_i\}$ such that

  $$|\{S \in \mathcal{F}_i : S \subseteq I\}| \geq (1 - 2^{-w})|\mathcal{F}_i|. \tag{3}$$
- If case (i) happens, we terminate the process. If case (ii) happens, set $U_{i+1} := I_i, \mathcal{F}_i := \{S \in \mathcal{F}_i : S \subset U_{i+1}\}$ and repeat the process.
- Terminate the process if $|\mathcal{F}_i| < 0.3|\mathcal{F}|$.

We make the following claims regarding the process:

Claim 3.16.

1. *If $\mathcal{F}_0 = \mathcal{F}$ is k-moonflower-free, then for all $i = 0, 1, 2, ...$ $\mathcal{F}_i$ is also k-moonflower-free.*
2. *$|\mathcal{F}_i| \geq 0.3|\mathcal{F}|$ for all $i \leq 2^w$.*
3. *For each i, $|U_{i+1}| = O(k \log(|U_i|/k) \log\log(|U_i|/k))$.*

Proof.

1. Suppose for contradiction that $\mathcal{F}_i$ contains a k-moonflower. Then since $\mathcal{F}_i \subseteq \mathcal{F}$, $\mathcal{F}$ must also contain a k-moonflower. This is a contradiction.
2. Iteratively applying (3), we know $|\mathcal{F}_i| \geq (1-2^{-w})^i|\mathcal{F}|$. The claim follows since $(1-1/x)^x \geq 0.3$ for all $x \geq 4$ and our assumption $w \geq 2$.
3. Use (1) and simplify.

□

Back to the process itself, if the process stops at some iteration j because of (i), then we have

$$0.3|\mathcal{F}| \leq |\mathcal{F}_j| \leq (Ck/w)^w, \tag{4}$$

where the first inequality holds by Claim 3.16(2).

Now consider the case where the process terminates as $|\mathcal{F}_i| < 0.3|\mathcal{F}|$. By Claim 3.16(2), we know $i_{\text{end}} \geq 2^w$. Let $g_i := |U_i|/k$ and notice that $g_0 \leq w$. By Claim 3.16(3),

$$g_{i+1} = O\Big(\log(|U_i|/k)\log\log(|U_i|/k)\Big) = O(\log g_i \cdot \log\log g_i).$$

This implies $g_i = O(1)$ whenever $i \geq \log^*(w)$. Consequently, at the end of the process we find,

$$|U_{i_{\text{end}}}| = O(k) \text{ and } |\mathcal{F}_{i_{\text{end}}}| \geq 0.3|\mathcal{F}|.$$

As a result,

$$|\mathcal{F}| \leq 4|\mathcal{F}_{i_{\text{end}}}| \leq 4\binom{|U_{i_{\text{end}}}|}{\leq w} \leq \left(\frac{Ck}{w}\right)^w$$

for some constant $C > 0$ as desired.

□

3.5 Iterated puncturing: $w \geq k$ case

Now we switch our focus to the case where $w \geq k$. We will prove the following theorem.

Theorem 3.17 (Iterated puncturing when $w \geq k$)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k-moonflower-free family of w-sets such that $w \geq k \geq 2$. Then $|\mathcal{F}| \leq (Cw/k)^k$ for some absolute constant C.*

While our overall strategy is similar to the $w \leq k$ case, we need to be more careful about the parameters we choose. One difficulty is the dependence on $\log(n/k)$ in the entropy bound. Removing unused elements from the universe if necessary, we may assume $n \leq (k-1)w$. In the $w \leq k$ case, $\log(n/k)$ is thus upper bounded by $O(\log(w)) = O(\log(\min\{k, w\}))$. However, in the $w \geq k$ case, $\log(n/k)$ is upper bounded by $O(\log(\max\{k, w\}))$, making it difficult to employ the same argument directly.

To resolve this issue, we will bring the dependence of the entropy bound on $\log(n/k)$ down to $\log(w/k)$. To achieve this, we employ an iterative universe reduction argument that replaces the universe size n by roughly $w \cdot \text{polylog}(n/k)$ while keeping at least a constant fraction of the family. As in the case when $w \leq k$, we first need an analogous one-step universe reduction lemma.

Lemma 3.18 (One-step universe reduction when $w \geq k$)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a k-moonflower-free family of w-sets such that $w \geq k$. Then either*

(i) $|\mathcal{F}| \leq 4t \cdot (4w/k)^{k/64}$ or

(ii) There exists a set $I \subseteq [n]$ of size $|I| \leq t$ such that

$$|\{S \in \mathcal{F} : S \subseteq I\}| \geq \frac{|\mathcal{F}|}{2}.$$

where $t := \min\{(k-1)w, O(w/p)\}$. Here $p \in (0,1)$ is chosen such that $h(n_0, k, w) = O(k\log(w/k))$ for $n_0 = (k-1)w$. In particular, we have

$$t \leq O\Big(w \cdot \psi\Big(\frac{\log(n/k)}{\log(w/k)}\Big)\Big)$$

where $\psi : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is defined as $\psi(x) = x\log x$.

Proof. We follow the same steps as in Lemma 3.15. By Lemma 2.9, we can truncate the universe to $U_0 = \mathrm{supp}(\mathcal{F})$ so that $n_0 := |U_0| \le (k-1)w$. By Lemma 3.13, we know $\mathcal{F}$ is $(p, 2^{h(n_0,k,p)})$-almost-covered where $h(n_0, k, p) := B \cdot k \log(n_0/k) \cdot \phi(p)$. Choose p such that

$$h(n_0, k, p) := B \cdot k \log(n_0/k) \cdot \phi(p) \le \frac{k \log(4w/k)}{64}.$$

We will justify why this is possible at the end of the proof. With this choice of p, we know $\mathcal{F}$ is (p, M)-almost-covered for $M := 2^{h(n_0,k,p)} \le (4w/k)^{k/64}$. Applying Theorem 3.12 with $\delta = 1/4$, there exists an $I \subseteq [n]$ with $|I| \le t := \min\{(k-1)w, \lceil \frac{2}{p}(w \ln 2 + \ln 4)\rceil\}$ such that

$$|\{S \in \mathcal{F} : S \subset I\}| \ge (1 - 1/4)|\mathcal{F}| - tM = \frac{3}{4}|\mathcal{F}| - tM.$$

Now if $tM \le \delta|\mathcal{F}|$, then

$$|\{S \in \mathcal{F} : S \subseteq I\}| \ge (1 - 2\delta)|\mathcal{F}| \ge \frac{|\mathcal{F}|}{2}$$

as desired. Otherwise, we have $tM > \delta|\mathcal{F}|$. Using $M \le (4w/k)^{k/64}$ and rearranging, we obtain the desired result.

Now we justified the choice of p. Let $p := a/(4 \log(4/a))$ where a is defined as

$$a := \frac{\log(4w/k)}{C \log(n_0/k)}$$

for some large enough constant C such that $a \in (0, 1/4]$. This allows us to apply Lemma 3.8 and obtain

$$h(n_0, k, p) := B \cdot k \log(n_0/k) \cdot \phi(p) \le B \cdot k \log(n_0/k) \cdot \frac{\log(4w/k)}{C \log(n_0/k)} = (B/C)k \log(4w/k).$$

Finally, we find

$$t \le \lceil \frac{2}{p}(w \ln 2 + \ln 4)\rceil \le O\Big(\frac{w}{p}\Big) = O\Big(\frac{w \log(1/a)}{a}\Big) \le O\Big(w \cdot \frac{\log(n/k)}{\log(w/k)} \cdot \log\Big(\frac{\log(n/k)}{\log(w/k)}\Big)\Big).$$

□

Given this, we follow the same strategy as in the proof of Theorem 3.14. By Lemma 2.9, we can truncate the universe to $U_0 = \mathrm{supp}(\mathcal{F})$ so that $n_0 := |U_0| \le (k-1)w$. We iteratively apply Lemma 3.18. At round r, we have a family $\mathcal{F}_r \subseteq 2^{U_r}$ with $|U_r| = n_r$ and $|\mathcal{F}_r| \ge 1$. Let $L_r := \log(en_r/k)$ be the entropy parameter. We stop if either (1) Lemma 3.18(i) happens or (2) L_r is small.

We first show that the parameters $\log(en_i/k)$ decreases logarithmically as a function of i.

Claim 3.19. *Let $L_r := \log(n_r/k)$ be defined as in the process above. Then*

$$L_{r+1} \le O\Big(\log(w/k) + \log L_r + \log\log L_r\Big)$$

In particular, if $L_r \ge C_1 \log(w/k)$ for some constant $C_1 > 0$, then $L_{r+1} \le L_r/2$.

Proof. By Lemma 3.18, we have

$$n_{r+1} \;\leq\; O\Big(w \cdot \psi\Big(\frac{\log(n_r/k)}{\log(w/k)}\Big)\Big)$$

Therefore,

$$\begin{aligned} L_{r+1} = \log\Big(\frac{n_{r+1}}{k}\Big) &\leq O\Big(\log(w/k) + \log\Big(\frac{\log(n_r/k)}{\log(w/k)}\Big) + \log\log\Big(\frac{\log(n_r/k)}{\log(w/k)}\Big)\Big) \\ &= O\Big(\log(w/k) + \log L_r + \log\log L_r\Big). \end{aligned}$$

For the halving statement, notice we would always have a logarithmic decrease unless $\log(w/k)$ is the dominate term. In which case, we have $L_{r+1} \leq O(\log(w/k))$ which is at most $L_r/2$ as long as $L_r \geq C_1 \log(w/k)$ for some large enough constant $C_1 > 0$. □

Using this, we are ready to prove the main theorem of this section.

Proof of Theorem 3.17. By Lemma 2.9, we can truncate the universe to $U_0 = \mathrm{supp}(\mathcal{F})$ so that $n_0 := |U_0| \leq (k-1)w$. We iteratively apply Lemma 3.18. At round r, we have a family $\mathcal{F}_r \subseteq 2^{U_r}$ with $|U_r| = n_r$ and $|\mathcal{F}_r| \geq 1$. Let $\beta := \log(4w/k)$ and $L_r := \log(n_r/k)$. We stop if either (1) Lemma 3.18(i) happens or (2) $L_r \leq C_1 \log(w/k)$ where C_1 is the constant from Claim 3.19.

First suppose we are always in the case of Lemma 3.18(ii). Then at each round, we obtain a subfamily $\mathcal{F}_{r+1} := \{S \in \mathcal{F}_r : S \subseteq I_{r+1}\}$ on universe $U_{r+1} := I_{r+1}$ with size $n_{r+1} := |U_{r+1}|$ satisfying $|\mathcal{F}_{r+1}| \geq |\mathcal{F}_r|/2$. By Claim 3.19, we always have $L_{r+1} \leq L_r/2$ while $L_r \geq C_1 \log(w/k)$. Therefore, since $n_0 = (k-1)w$, after at most

$$R \leq 1 + \Big\lceil \log\Big(\frac{L_0}{C_1 \log(w/k)}\Big)\Big\rceil = O(\log L_0) = O(\log\log(w))$$

rounds, we reach a round with $L_R \leq C_1 \log(w/k)$. Applying Lemma 3.18 one more time, we have

$$|U_{R+1}| = O\left(w \cdot \psi\left(\frac{L_R}{\log(w/k)}\right)\right) = O(w).$$

The claim now follows: Applying Lemma 3.6 to $\mathcal{F}_{R+1}$ which has VC dimension at most $k-1$, we have

$$|\mathcal{F}_{R+1}| \leq \sum_{i=0}^{k-1} \binom{|U_{R+1}|}{i} \leq \left(\frac{Cw}{k}\right)^k$$

and hence

$$|\mathcal{F}| \leq 2^{R+1} \cdot |\mathcal{F}_{R+1}| \leq \left(\frac{C'w}{k}\right)^k$$

where C, C' are some absolute constants.

Lastly, if our process terminates because Lemma 3.18(i) happens at some $r \leq R+1$, then

$$|\mathcal{F}_r| \leq 4t_r \cdot (4w/k)^{k/64}.$$

Since $|\mathcal{F}_r| \geq |\mathcal{F}_R|$, we thus have

$$|\mathcal{F}| \leq 2^{R+1}|\mathcal{F}_r| \leq O(\log w) \cdot 4t_r \cdot (4w/k)^{k/64} \leq \Big(\frac{C''w}{k}\Big)^k$$

for some large enough constant C''. The last inequality holds since $t_r \leq (k-1)w$ for all r. □

4 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 $\mathsf{poly}(\log(k/\varepsilon), \log\log n)$.

Throughout, let $\mathcal{C} \subseteq \{0,1\}^n$ be a code with $\mathrm{NRD}(\mathcal{C}) \le k-1$. Equivalently, the support family $\mathsf{Supp}(\mathcal{C}) := \{\mathrm{supp}(x) : x \in \mathcal{C}\}$ is k-moonflower-free by Lemma 2.12.

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

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

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

Theorem 4.1 (Improved sparsification with a single $\log n$ factor)**.** *Let $\mathcal{C} \subseteq \{0,1\}^n$ satisfy $\mathrm{NRD}(\mathcal{C}) \le k-1$. Then for every $\varepsilon \in (0, 1/4)$ there exists a weighted coordinate set (T, α) that ε-sparsifies $\mathcal{C}$ and satisfies*

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

The rest of this section is organized as follows: in Section 4.1, we state and prove the precise puncturing result we need for sparsification. Given this, in Section 4.2, we give a full proof of Theorem 4.1.

4.1 Puncturing lemma for sparsification

We first state the main lemma of this subsection which is going to play an important role in the proof of Theorem 4.1.

Lemma 4.2 (Simplified weight-scale puncturing bound for sparsification layers)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be k-moonflower-free family, where all sets $S \in \mathcal{F}$ have sizes between w and $2w$. Fix parameters $\eta \in (0, 1/4)$ and $\theta \in (0,1)$. Then there exists $I \subseteq [n]$ such that*

$$|\mathcal{F}_{\overline{I}}| \;\le\; |I| \cdot \exp(\theta \eta^2 w),$$

and

$$|I| \;\le\; \frac{k}{\theta\, \eta^2} \cdot \mathsf{poly}(\log(kw/\eta\theta))\,.$$

To prove Lemma 4.2, we need the following version of the one-step puncturing lemma.

Corollary 4.3 (Trace puncturing of good families)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of w-sets. Fix $p \in (0,1)$ and $M \ge 1$. If $\mathcal{F}$ is (p, M)-almost-covered, then there exists $I \subseteq [n]$ with*

$$|I| \;\le\; t \;:=\; \min\left\{n,\; \left\lceil \frac{2}{p} \ln\left(|\mathcal{F}| \cdot 2^w\right) \right\rceil \right\}$$

such that

$$|\mathcal{F}_{\overline{I}}| \;\le\; t \cdot M.$$

Proof. This is exactly the potential argument used in Theorem 3.12, run until the trace family becomes empty. $\square$

Now we apply Corollary 4.3 with the appropriate parameters to obtain the following Lemma.

Lemma 4.4 (Weight-scale puncturing with tunable trace bound). *There exists an absolute constant $c_0 \ge 1$ such that the following holds. Let $k \ge 2$ and let $\mathcal{F} \subseteq 2^{[n]}$ be a k-moonflower-free family of w-sets. Fix parameters $\eta \in (0, 1/4)$ and $\theta \in (0, 1)$. Then there exists $I \subseteq [n]$ such that*

$$\begin{aligned} |I| &\le c_0 \cdot \frac{k \log(w)}{\theta\,\eta^2} \cdot \left(1 + \frac{\log|\mathcal{F}|}{w}\right) \cdot \max\left(1, \log\left(\frac{k \log(w)}{\theta\,\eta^2 w}\right)\right), \\ |\mathcal{F}_{\overline{I}}| &\le |I| \cdot \exp(\theta\,\eta^2 w). \end{aligned}$$

Proof. Let $n_0 := |\mathrm{supp}(\mathcal{F})|$. Since $\mathcal{F}$ is k-moonflower-free and all sets have size $\le w$, Lemma 2.9 implies $n_0 \le kw$, hence $\log(en_0/k) \le O(\log(w))$.

Apply Lemma 3.13 to $\mathcal{F}$ with parameter k and a value of $p \in (0, 1)$ to be chosen, obtaining that $\mathcal{F}$ is (p, M)-almost-covered with

$$M \le 2^{h(n_0,k,p)}, \qquad h(n_0, k, p) = B\,k \log\Big(\frac{n_0}{k}\Big) \cdot p \log(1/p).$$

Choose

$$a := \min\left\{\frac{1}{4}, \frac{\theta\,\eta^2 w}{B\,k \log(\frac{n_0}{k})}\right\}, \qquad p := \frac{a}{4 \log(4/a)}.$$

By Lemma 3.8, $p \log(1/p) \le a$, hence $h(n_0, k, p) \le \theta\,\eta^2 w$, and therefore

$$M \le 2^{\theta\eta^2 w} \le \exp(\theta\eta^2 w).$$

Now apply Corollary 4.3 to obtain I with

$$|I| \le O\Big(\frac{w + \log|\mathcal{F}|}{p}\Big).$$

Using $1/p = O(\log(1/a)/a)$ and $a = O(\theta\eta^2 w/(k \log(w)))$ gives the stated bound on $|I|$, after absorbing constants and using $\log(en_0/k) \le O(\log(w))$. Finally, $|\mathcal{F}_{\overline{I}}| \le |I| \cdot M \le |I| \exp(\theta\eta^2 w)$. □

It remains to upper bound the quantity $\log|\mathcal{F}|/w$.

Lemma 4.5 (Layer-size bound for moonflower-free families). *Let $\mathcal{F} \subseteq 2^{[n]}$ be k-moonflower-free of w-sets. Then there is an absolute constant $C_{\mathrm{lay}} \ge 1$ such that*

$$\frac{\log|\mathcal{F}|}{w} \le C_{\mathrm{lay}} \cdot \begin{cases} 1 + \log\Big(\dfrac{k}{w}\Big) & \text{if } w \le k, \\ 1 & \text{if } w \ge k. \end{cases}$$

Proof. Theorem 3.1 gives that there is an absolute constant C such that

$$|\mathcal{F}| \le \begin{cases} \left(\dfrac{Ck}{w}\right)^w & \text{if } w \le k, \\ \left(\dfrac{Cw}{k}\right)^k & \text{if } w \ge k. \end{cases}$$

This implies that

$$\frac{\log|\mathcal{F}|}{w} \le \begin{cases} \log\left(\dfrac{Ck}{w}\right) & \text{if } w \le k, \\ \frac{k}{w} \log\left(\dfrac{Cw}{k}\right) & \text{if } w \ge k. \end{cases}$$

The bound in the case of $w \le k$ is $\log(C) + \log(k/w)$, and in the case of $w \ge k$ is bounded by $\log(C) + 1/2$, since $x\log(1/x) < 1/2$ for all $x \in (0,1)$.

□

Proof of Lemma 4.2. Start from Lemma 4.4:

$$|I| \le c_0 \cdot \frac{k\log(w)}{\theta\,\eta^2} \cdot \left(1 + \frac{\log|\mathcal{F}|}{w}\right) \cdot \log\left(\frac{k\log(w)}{\theta\,\eta^2 w}\right).$$

Apply Lemma 4.5 to bound $\frac{\log|\mathcal{F}|}{w} \le O(\log k)$. This gives

$$|I| = \frac{k}{\theta\eta^2}\mathsf{poly}(\log k, \log w, \log \eta^{-1}, \log\theta^{-1}) = \frac{k}{\theta\,\eta^2} \cdot \mathsf{poly}(\log(kw/\eta\theta)).$$

□

4.2 Proof of Theorem 4.1

The proof of Theorem 4.1 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 $\frac{k}{\varepsilon^2} \cdot \mathsf{poly}(\log(k/\varepsilon), \log\log n)$. Since there are $R = \Theta(\log n)$ rounds, the final puncturing set has size $|T| \le \frac{k\log n}{\varepsilon^2} \cdot \mathsf{poly}(\log(k/\varepsilon), \log\log n)$. The remaining coordinates will be small due to halving in each round.

- The puncturing is delicate in that for each weight range w, we choose a specific puncturing set $I_{r,w}$. For round r, let $\mathcal{F}^{(r)}_{(w,2w]}$ 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 coordiantes 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 $1/\mathsf{poly}(\log n)$ 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 $\mathsf{poly}(\log n)$ 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 $w_\star \approx O((k\log(k/\epsilon) + \log\log n))/\epsilon^2$ and do the following. Pick a dyadic weight $w = 2^i w_{min} > w_\star$. For such codewords, we do no puncturing, and instead union-bound directly over such codewords using our improved 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 succedinng 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, $1 \pm \epsilon_r$, incurred by the sampling process satisfies

$$\epsilon_r := \begin{cases} 0 & w_r \leq w_{min}, \\ \eta_0 & w_{min} < w_r \leq w_\star \\ \eta(w) & w_r > w_\star \end{cases}$$

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

  There can be several rounds where the codeword would be large weight (i.e, weight above $w_\star$). 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 4.1) Fix $R := \lceil \log_2 n \rceil$ and sampling rate $q := 1/2$.

Parameter choices. In what follows, let C be a large constant. Define the transition threshold

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

Define the medium-weight per-round error

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

Define the tiny-weight cutoff

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

Finally, for dyadic $w > w_\star$ 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 $r = 0, 1, \ldots, R-1$, 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 dyadic[4] $w \in [w_{\min}, w_\star]$, define

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

Define the tiny family

$$\mathcal{F}^{(r)}_{\leq w_{\min}} := \Big\{ \operatorname{supp}(x) \cap U_r : \; x \in \mathcal{C}, \quad |\operatorname{supp}(x) \cap U_r| \leq w_{\min} \Big\}.$$

[4] Here by dyadic, we mean $w = 2^j w_{\min}$ for $j = 0, 1, ..., \log(w_\star / w_{\min})$.

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

(a) Tiny weights: capture deterministically. Let

$$I_{r,\mathrm{tiny}} := \mathrm{supp}\Big(\mathcal{F}^{(r)}_{\le w_{\min}}\Big) \subseteq U_r.$$

By Lemma 2.9, $|I_{r,\mathrm{tiny}}| \le k\, w_{\min}$.

(b) Medium weights: puncture to shrink traces. For each dyadic $w \in [w_{\min}, w_\star]$, apply Lemma 4.4 to $\mathcal{F}^{(r)}_{(w,2w]}$ with parameters $\eta = \eta_0$ and $\theta := 1/100$ to obtain $I_{r,w} \subseteq U_r$ such that

$$\Big|\Big(\mathcal{F}^{(r)}_{(w,2w]}\Big)_{\overline{I_{r,w}}}\Big| \;\le\; |I_{r,w}| \cdot \exp\Big(\frac{1}{100}\eta_0^2 w\Big)\,. \tag{5}$$

(c) Define I_r and recurse. Let

$$I_r \;:=\; I_{r,\mathrm{tiny}} \;\cup \bigcup_{\substack{\text{dyadic } w\\ w_{\min}\le w\le w_\star}} I_{r,w}, \qquad V_r := U_r \setminus I_r,$$

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

(d) Output. Define

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

(The sets $I_0, \ldots, I_{R-1}, U_R$ 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 $w \in [w_{\min}, w_\star]$. Consider the trace family on V_r:

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

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

$$\Big|\Big(\mathcal{F}^{(r)}_{(w,2w]}\Big)_{V_r}\Big| \le \Big|\Big(\mathcal{F}^{(r)}_{(w,2w]}\Big)_{\overline{I_{r,w}}}\Big|.$$

Moreover, for any A in this family, we have $|A| \le 2w$. Applying Lemma 2.13 with $\Delta := \eta_0 w$ gives

$$\mathbf{Pr}\Big[\big|\,2|A \cap U_{r+1}| - |A|\,\big| > \eta_0 w\Big] \le 2\exp\big(-\Omega(\eta_0^2 w)\big)\,.$$

Using (5) and the definition of $w_{\min}$ (so $\eta_0^2 w \gg \log|I_{r,w}|$), the union bound over all $A \in (\mathcal{F}^{(r)}_{(w,2w]})_{V_r}$ fails with probability at most $\exp(-\Omega(\eta_0^2 w)) \le 1/(100R\log(2w_\star))$ after increasing constants.

Large regime. Fix dyadic $w > w_\star$ and consider the trace family

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

Since $\mathcal{F}^{(r)}_{(w,2w]}$ is k-moonflower-free and $w \geq k$ in this regime, Theorem 3.1 gives

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

For any $A \in \mathcal{H}^{(r)}_{(w,2w]}$ we have $|A| \leq 2w$, so by Lemma 2.13 with $\Delta := \eta(w)\, w$,

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

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

Combining medium and large regimes and summing over the $O(\log w_\star)$ medium scales, we obtain that conditioned on the past, round r fails with probability at most $1/(50R)$. A union bound over $r = 0, \ldots, R-1$ yields that all rounds succeed with probability at least $49/50$. Also $\mathbb{E}[|U_R|] \leq n2^{-R} \leq 1$, hence $\mathbf{Pr}[|U_R| \leq 50] \geq 49/50$ 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 $S := \mathrm{supp}(x)$.

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

Consider the next weight estimate

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

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

$$\epsilon_r := \begin{cases} 0 & w_r \leq w_{min}, \\ \eta_0 & w_{min} < w_r \leq w_\star \\ \eta(w) & w_r > w_\star \end{cases} .$$

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

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

Thus,

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

Combining the above equations, we get

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

Thus, we have

$$(1 - \epsilon_r)\widehat{N}_r \leq \widehat{N}_{r+1} \leq (1 + \epsilon_r)\widehat{N_r}.$$

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

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

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

$$\widehat{N}_0 \cdot \exp\left(-2\sum_{r=0}^{R-1} \epsilon_r\right) \leq \widehat{N}_R \leq \widehat{N}_0 \cdot \exp\left(\sum_{r=0}^{R-1} \epsilon_r\right). \tag{6}$$

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

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

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

by the definition of $w_\star$ (with C large enough).

Combining the above we get that $\sum_r \epsilon_r \leq \varepsilon/25$. Thus, in particular we must have $\widehat{N}_R = (1 \pm \varepsilon)N_0$, 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 $|I_{r,\text{tiny}}| \leq kw_{\min}$.

For each dyadic $w \in [w_{\min}, w_\star]$, apply Lemma 4.2 to the layer family $\mathcal{F}^{(r)}_{(w,2w]}$ with $\eta = \eta_0$ and $\theta = 1/100$ to obtain

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

Since $w \leq w_\star$ throughout, we have $\log w \leq \log w_\star$ and

$$\log(k/\eta_0) = \log(k/\varepsilon) + O(\log\log w_\star),$$

hence

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

using $\log w_\star = O(\log(k/\varepsilon) + \log\log n)$ by the definition of $w_\star$.

There are $O(\log(2w_\star))$ such dyadic values. Thus

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

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

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

as claimed.

This finishes the proof of the full sparsification theorem. □

4.3 Lower bound

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

Lemma 4.6 (Lemma 1.9, 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 $\mathrm{NRD}(\mathcal{C}) = k$ 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 $a_1 \leq \cdots \leq a_s$ be a maximal collection of integers such that $1 \leq a_1$, $a_s < m$ and $a_{j+1} > (1+\varepsilon)a_j$ 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 $[i : j] = \{i, i+1, \ldots, j\}$. 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 $\mathrm{NRD}(\mathcal{C}_i) = 1$ and hence [5] $\mathrm{NRD}(\mathcal{C}) = k$. 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 $|T \cap \mathsf{Supp}(\mathcal{C}_i)| < |\mathcal{C}|/k = s$. 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 $\mathcal{C}$.

□

[5] A 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.

[CKR22] Bruno Pasqualotto Cavalar, Mrinal Kumar, and Benjamin Rossman. Monotone circuit lower bounds from robust sunflowers. *Algorithmica*, 84(12):3655–3685, 2022.

[ER60] P. Erdős and R. Rado. Intersection theorems for systems of sets. *Journal of the London Mathematical Society*, 35(1):85–90, 1960.

[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.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il