Electronic Colloquium on Computational Complexity, Report No. 192 (2025)

Nearly Tight Lower Bounds for Relaxed Locally Decodable Codes via Robust Daisies

Guy Goldberg *
Weizmann Institute of Science
guy.goldberg@weizmann.ac.il

Tom Gur†
University of Cambridge
tom.gur@cl.cam.ac.uk

Sidhant Saraogi‡
Georgetown University
ss4456@georgetown.edu

November 24, 2025

Abstract

We show a nearly optimal lower bound on the length of linear relaxed locally decodable codes (RLDCs). Specifically, we prove that any q-query linear RLDC $C\colon \{0,1\}^k \to \{0,1\}^n$ must satisfy $n = k^{1+\Omega(1/q)}$. This bound closely matches the known upper bound of $n = k^{1+O(1/q)}$ by Ben-Sasson, Goldreich, Harsha, Sudan, and Vadhan (STOC 2004).

Our proof introduces the notion of robust daisies, which are relaxed sunflowers with pseudorandom structure, and leverages a new spread lemma to extract dense robust daisies from arbitrary distributions.

*GG is supported by ERC Consolidator Grant 772839 and ISF Grant 2073/21.
†TG is supported by ERC Starting Grant 101163189 and UKRI Future Leaders Fellowship MR/X023583/1.
‡SS is supported by the NSF CAREER grant CCF-1845125.

ISSN 1433-8092

1 Introduction

In their influential 2004 paper, Ben-Sasson, Goldreich, Harsha, Sudan, and Vadhan (BGHSV) [BGH+06] introduced the notion of relaxed locally decodable codes (RLDCs). Similarly to standard locally decodable codes (LDCs), these are error-correcting codes from which individual message bits can be recovered, with high probability, by querying only a few codeword bits, even when the codeword is partially corrupted. However, RLDCs permit a relaxed decoder that, on a small fraction of coordinates, may output the rejection symbol $\perp$ upon detecting corruption.

More precisely, a (q, δ, σ)-RLDC $C : \{0,1\}^k \to \Sigma^n$ is a code that admits a relaxed decoder D that, given an index $i \in [k]$ and oracle access to $w \in \{0,1\}^n$ that is δ-close to some codeword $c = C(x)$, satisfies the following conditions.

1. *Completeness:* if $w = c$, then $D^w(i) = x_i$.
2. *Relaxed local decoding:* otherwise, $\Pr[D^w(i) \in \{x_i, \perp\}] \geq \sigma$.

As observed in [BGH+06, Lemma 4.10], for $O(1)$-query RLDCs, the two conditions above imply a relaxed decoder that will only output $\perp$ on an arbitrarily small fraction of the message bits.

This seemingly modest relaxation of LDCs allows for constructions with dramatically better parameters. BGHSV constructed q-query linear RLDCs with length $n = k^{1+O(1/q)}$.[1] In particular, this implies $O(1)$-query RLDCs with nearly-linear length, while the best known construction of $O(1)$-query (non-relaxed) LDCs has superpolynomial length [Yek08, Efr12]. However, despite the much attention that RLDCs received (cf. [GR18, GG18, GGK19, GRR20, GL21, AS21, GG21, DGMT22, CGS22, CY22, DGL23, Gol24b, KM24, CY24]), there are no constructions that improve on BGHSV by achieving length $n = k^{1+o(1/q)}$, and whether such constructions are possible remained an open problem.

1.1 Main result

We prove a lower bound for linear RLDCs, which closely matches the $n = k^{1+O(1/q)}$ upper bound of BGHSV. We do this by first proving a lower bound for RLDCs with *non-adaptive* decoders, and then apply a known reduction by Goldberg [Gol24a] to deduce the same bound for linear RLDCs.

Theorem 1. *Let $C : \{0,1\}^k \to \Sigma^n$ be a non-adaptive (q, δ, σ)-RLDC, where $q \in \mathbb{N}$, $\sigma > 0$, and $\delta > n^{-\frac{\sigma}{2q}}$. Then,*

$$n \geq \left(\frac{\sigma^2 \cdot k}{38q^4 \log^2(|\Sigma|) \cdot \log^2 k} \right)^{1+\frac{1}{\lceil q/\sigma \rceil}}.$$

For simplicity, throughout the rest of this section, we restrict our attention to the standard setting of $\sigma = 2/3$ and $\delta = \Omega(1)$, and assume a binary alphabet $\Sigma = \{0,1\}$. In this setting, and for a constant q, Theorem 1, yields the following.

Corollary 2. *For any linear q-RLDC $C : \{0,1\}^k \to \{0,1\}^n$ it holds that $n = k^{1+\Omega(1/q)}$.*

[1] While BGHSV only guaranteed that the length is $n = k^{1+O(1/\sqrt{q})}$, Goldreich [Gol24b] showed that their construction achieves the stronger guarantee, with minor modifications to the analysis.

Our result improves upon the previous state-of-the-art lower bound of $n = k^{1+\Omega(1/q^2)}$ for linear RLDCs. This prior bound was achieved by applying Goldberg's reduction [Gol24a] to the $n = k^{1+\Omega(1/q^2)}$ non-adaptive bound in [Gol23] (which improved on the $n = k^{1+\Omega(1/q^2 \log^2 q)}$ bound in [GL21]).

For the constant rate regime, rearranging the terms in Theorem 1 yields the following.

Corollary 3. *For any linear q-RLDC $C : \{0,1\}^k \to \{0,1\}^n$, where $n = O(k)$, it holds that $q = \Omega\left(\frac{\log k}{\log \log k}\right)$.*

Since linear locally correctable codes (RLCCs) imply RLDCs with the same parameters, both corollaries immediately extend to them.

Beyond non-adaptive RLDCs. We remark that the machinery developed in [DGL23] provided the means to extend the $n = k^{1+\Omega(1/q^2 \log^2 q)}$ lower bound in [GL21] to (not necessarily linear) adaptive RLDCs. We believe that an adaptation of this machinery will extend our lower bounds to general RLDCs. We leave this to future work.

1.2 Robust daisies

The proof of our main result, Theorem 1, introduces the combinatorial notion of *robust daisies*, which may be of independent interest. A distribution over sets is a robust daisy with a kernel K if, after removing the kernel, its support forms a *satisfying set system*. That is, a binomial sampling of the universe elements contains a full set from this system (ignoring the elements in K) with high probability (see Definition 4.1). Furthermore, this property must hold for any subset of the support, where the required success probability scales exponentially in the subset's density.

Definition 1.1. *(Robust daisy) A distribution μ over $\mathcal{P}(U)$ is a (p, ε)-*robust daisy *with kernel $K \subseteq U$, if, for every $\mathcal{D} \subseteq \mathrm{supp}(\mu)$:*

$$\Pr_{W \sim Bin(U,p)}[\exists\, S \in \mathcal{D}, S \subseteq K \cup W] \geq 1 - \varepsilon^{\mu(\mathcal{D})}\,.$$

For any set $S \in \mathrm{supp}(\mu)$, we call $S \setminus K$ a *petal* of the robust daisy. Put differently, the robust daisy condition dictates that every subset of petals $\{S \setminus K \mid S \in \mathcal{D}\}$ is $\left(p, \varepsilon^{\mu(\mathcal{D})}\right)$-satisfying.

Robust daisies are closely related to robust sunflowers [Ros14]. They both require the petals to be a satisfying set system. However, the notions differ in two ways: (1) for robust daisies, the kernel is allowed to have an arbitrary structure, rather than being restricted to the intersection of all sets; (2) the robust daisy is a distribution over sets, rather than an unweighted set system, and the satisfying set system condition must hold not only for the support of the distribution, but also for its subsets.

We remark that, unlike the notion of (non-robust) daisies [GL21, DGL23], where outside the relaxed kernel each point is only required to be covered by a bounded number of sets, robust daises capture a pseudorandom structure, known as *spreadness*, outside of the kernel. Indeed, to argue about robust daisies, we prove a new spread lemma that is applicable to spread families of small sets (see Section 4).

Robust daisy lemma. Our main structural result concerning robust daisies shows that it is always possible to extract a dense robust daisy from a distribution over small sets.

Lemma 1.2. *(The Robust Daisy Lemma; informally stated, see Lemma 5.1) Fix $q \in \mathbb{N}$ and a set U of size n. Let μ be a distribution over $\mathcal{P}(U)$ such that $|S| \leq q$ for every $S \in \mathrm{supp}(\mu)$. Then, there exists $\mathcal{D} \subseteq \mathrm{supp}(\mu)$ with $\mu(\mathcal{D}) \geq 0.99$ and a kernel $K \subseteq U$ with $|K| = o(n)$ such that the conditional distribution $\mu_{\mathcal{D}}(x) = \frac{\mu(x)}{\mu(\mathcal{D})}$ is a (p, ε)-robust daisy with kernel K, where*

$$p = n^{-\Theta(1/q)} \quad \text{and} \quad \varepsilon = 2^{-\Omega(|K|)} .$$

The conceptual message of the robust daisy lemma stands in sharp contrast to that of the robust sunflower lemma. While the robust sunflower lemma shows the existence of a small, highly structured subset within the set system, the robust daisy lemma instead extracts a pseudorandom approximation of the entire distribution over the set system. To use a metaphor, if a robust sunflower is a small precious "gem" that can be found within any large enough mountain, a robust daisy is (approximately) *the entire mountain*.

1.3 Related work

LDCs, LCCs and their relaxed counterparts have attracted significant attention in recent years. See the works of Yekhanin [Yek12] and Kopparty and Saraf [KS17] and references therein for comprehensive surveys of LDCs, LCCs and their applications.

RLDCs constructions. The constructions of RLDCs and RLCCs can be separated into two main parameter regimes: constant query complexity, and constant rate.

In the constant rate regime, the state-of-the-art code is the construction by Cohen and Yankovitz [CY24]. They construct a linear RLCC with rate arbitrarily close to 1, and query complexity $q = (\log n)^{2+o(1)}$. This construction builds upon the result by Kumar and Mon [KM24], which shows a similar code but with query complexity $q = (\log n)^{O(1)}$.

In the constant query regime, the original work of [BGH+06] claimed to achieve RLDC with constant query complexity $O(q)$ and length $n = O(k^{1+1/\sqrt{q}})$. In fact, [Gol24b] showed that this construction actually achieves $n = O(k^{1+1/q})$, which still makes it the current state-of-the art RLDC construction with constant query complexity.

The work of [GRR20] introduced the notion of RLCCs, constructing such a code with constant query complexity, but with a worse length tradeoff. Chiesa, Gur, and Shinkar [CGS22] constructed an improved RLCC, achieving length $n = O(k^{1+1/\sqrt{q}})$ (matching the original BGHSV claim). This was later improved by Asadi and Shinkar [AS21], who constructed an RLCC with length $n = O(k^{1+1/q})$, matching the actual (and stronger) bound of the BGHSV construction.

Lower bounds. In recent decades, extensive research has been conducted on lower bounds for (non-relaxed) LDCs in various regimes [KT00, KdW03, Woo07, Woo12, AGKM22, JM25, BHKL25].

Gur and Lachish [GL21] presented the first lower bound for relaxed LDCs. Specifically, they showed that any non-adaptive RLDC requires a block length of $n = k^{1+\Omega\left(\frac{1}{q^2 \log^2 q}\right)}$. For the adaptive case, they established a lower bound of $n = k^{1+\Omega\left(\frac{1}{2^{2q} \log^2 q}\right)}$.

The result of [GL21] was extended to additional settings, such as proofs of proximity and property testing, and to the adaptive setting by Dall'Agnol, Gur and Lachish [DGL23]. Specifically, they extended the lower bound of $n = k^{1+\Omega\left(\frac{1}{q^2 \log^2 q}\right)}$ to *adaptive* RLDCs.

Goldreich [Gol23] surveyed and simplified the work of [GL21], without employing the newer techniques of [DGL23]. He established an improved bound of $n = k^{1+\Omega(1/q^2)}$ for the non-adaptive case, and a bound of $n = k^{1+\Omega(1/q^3)}$ for the adaptive case (which is weaker than the one presented in [DGL23]).

Goldberg [Gol24a] presented a generic reduction that transforms any lower bound for non-adaptive RLDCs and extends it to (possibly adaptive) linear RLDCs. Applying this reduction to the bound from [Gol23] extends the $n = k^{1+\Omega(1/q^2)}$ lower bound to all linear RLDCs.

Spreadness. Our techniques draw on the powerful connection between spreadness and robust combinatorial structures, a link that has been central to recent breakthroughs.

The concept of spreadness for distributions was introduced by Talagrand [Tal10]. A version of the spread lemma, with roots in Rossman [Ros14], was famously used by Alweiss, Lovett, Wu, and Zhang [ALWZ21] (building on [LSZ20]) to prove that any sufficiently spread set system contains a robust sunflower. This directly led to a breakthrough on the sunflower lemma. This line of work, and the spread lemma itself, has since been significantly strengthened [FKNP21, PP23, Rao20] and has found numerous applications across combinatorics and computer science (among others, see [ALWZ21, FKNP21, PP23, CKR22, CGR+25]). For a detailed survey, we refer the reader to [Rao25].

1.4 Open problem

Our work leaves several interesting directions for further research. We highlight one open question that we find particularly compelling.

In the constant rate regime, where $n = O(k)$, Theorem 1 implies a lower bound on the query complexity, of $q = \Omega(\frac{\log k}{\log \log k})$. On the other hand, the recent state-of-the-art construction of a constant-rate RLDC by Cohen and Yankovitz [CY24], achieves $q = O(\log^2 k)$. An important open problem is to close the quadratic gap that still remains in this regime.

2 Proof overview

The proof of the RLDC lower bound in Theorem 1 consists of the following three high-level steps.

- **Step 1: Reduction to a combinatorial problem.** First, we reduce the problem of proving a lower bound for RLDCs to a purely combinatorial problem: finding a specific structure, a dense robust daisy, within the decoder's query-set distribution.
- **Step 2: From spreadness to robust daisies.** Second, we establish the key link between spreadness and our new notion of robust daisies. We introduce a set-theoretic property, which is a generalization of the well-known notion of *set spreadness*. We prove a new *small-set spread lemma* which shows that any spread set system is *satisfying*, which is the required structure of the robust daisy outside the kernel.

- **Step 3: finding spreadness.** Third, we prove a *spreadness extraction lemma*. We show that every distribution over sets can be made spread, by *puncturing* (removing $o(n)$ elements from the universe), and *conditioning* (restricting the distribution to a large-measure subset of its support).

These three components chain together to prove the main theorem. We apply the spreadness extraction lemma (Step 3) to the RLDC's query-set distribution to find a large, spread substructure. Our small-set spread lemma (Step 2) then proves this structure is a robust daisy. Finally, by our reduction (Step 1), the existence of this robust daisy within the decoder's queries implies the $n = k^{1+\Omega(1/q)}$ lower bound. We proceed to elaborate on each of these three steps.

2.1 RLDC lower bounds via robust daisies

Our first contribution is conceptual: we abstract and generalize the argument underlying the RLDC lower bound in [GL21]. This abstraction is crucial, as it reveals the bottleneck common to all previous lower bounds, including [DGL23, Gol23, Gol24a], and provides the generality that is needed to surpass the barrier of $n = k^{1+\Omega(1/q^2)}$ lower bounds.

In the following, let $C : \{0,1\}^k \to \{0,1\}^n$ be a non-adaptive q-RLDC; that is, for each decoding index $i \in [k]$ the decoder's queries are determined by a query-set distribution μ_i over q-tuples of codeword coordinates.

We show that if the relaxed decoder is structured in the sense that each distribution μ_i constitutes a robust daisy (Definition 1.1), then the following lower bound on the code's block length must hold.

Lemma 2.1. *(informally stated, see Lemma 6.3) Let $C : \{0,1\}^k \to \{0,1\}^n$ be a non-adaptive q-query RLDC. If for every $i \in [k]$, the query-set distribution μ_i is a (p, ε_i)-robust daisy with a kernel $K_i \subseteq [n]$ such that $|K_i| = o(n)$ and $\varepsilon_i = 2^{-\Omega(|K_i|)}$, then $k \leq pn$.*

We defer a detailed overview of the proof of Lemma 2.1 to Section 6.1, which we encourage readers unfamiliar with the techniques of [GL21] to review first. Our focus here is on how the reduction to a combinatorial problem, which Lemma 2.1 provides, allows us to overcome the limitations of previous approaches.

Towards this end, note that the query-set distribution μ of a general RLDC might not form a robust daisy, but rather an arbitrary set of distributions $\{\mu_i\}_{i\in[k]}$ supported on q-sets. However, to apply the reduction, it suffices that each μ_i can be *approximated* by a robust daisy.

To see this, fix $i \in [k]$ and denote $\mu = \mu_i$. Note that if there exists a dense sub-family $\mathcal{D} \subseteq \mathrm{supp}(\mu)$ such that the conditional distribution $\mu_{\mathcal{D}}$ is a robust daisy, we can slightly modify the operation of the relaxed decoder: instead of sampling a set $S \sim \mu$ (as the original decoder does), the modified decoder will sample a set $S \sim \mu_{\mathcal{D}}$. Hence, if $\mathcal{D}$ is dense enough (say, with $\mu(\mathcal{D}) \geq 0.99$), then the soundness probability of the modified decoder is only slightly worse than that of the original one.

The above discussion, combined with Lemma 2.1, reduces the task of proving RLDC lower bounds to the following, purely combinatorial problem.

Problem (Robust daisy extraction)**.** *Given an arbitrary distribution μ supported on q-sets, extract a dense (p, ε)-robust daisy with kernel K such that $\varepsilon = 2^{-\Omega(|K|)}$, while minimizing p.*

Our reduction to the problem of extracting robust daisies, which are closely related to robust sunflowers (see Section 1.2), exposes the connection to the pseudorandom structure captured by *spread lemmas*, which lies at the heart of robust sunflower lemmas and is also a key component in our proof.

We remark that previous RLDC lower bounds did not employ an abstract reduction; rather, they directly analyzed the query-set distribution to identify specific structures. Recasting the works of [GL21, DGL23, Gol23] through the lens of robust daisy extraction, the structures they isolate are in fact (p, ε)-robust daisies, albeit with relatively weak parameters: to achieve ε that is exponentially small in the kernel's size, they need $p = n^{-\Theta(1/q^2)}$. The quadratic dependency in the query complexity gives the corresponding factor in the lower bound.[2] Moreover, as we shall see next, there are explicit counterexamples showing that these structures cannot attain the parameters needed to surpass $n = k^{1+\Omega(1/q^2)}$ lower bounds.

2.2 Extracting a robust daisy

In light of the reduction above, we can set aside RLDCs and focus on the combinatorial problem of robust daisy extraction. For simplicity, let us denote a (p, ε)-robust daisy with $\varepsilon = 2^{-\Omega(|K|)}$ as a *p-robust daisy*.

We begin by examining the methods used in previous lower bounds to extract structures that can be viewed as p-robust daisies, and explain a barrier for such approaches.

The bottleneck: t-daisies. The combinatorial structure extracted in all previous works [GL21, DGL23, Gol23] is that of a *t-daisy*.[3] A set system is a t-daisy if, outside a kernel K, each element is contained in at most t sets.

Reframing [GL21] through the abstraction of robust daises, their argument regarding t-daisies can be seen as showing for any set system containing $O(n)$ sets, the following holds:

1. It is always possible to extract a t-daisy with $t|K| = O(n^{1-1/q})$.[4]
2. If a t-daisy satisfies the condition above, then it is a $n^{-\Theta(1/q^2)}$-robust daisy.

We remark that the second item is shown by simple, first-principle arguments. Namely, using a greedy process, [GL21] finds a family of disjoint sets in the t-daisy. Then, they directly calculate the probability to sample one of these sets.

We argue that $p = n^{-\Theta(1/q^2)}$ is the best (i.e., minimal) sampling probability that can be achieved by extracting t-daisies; hence, they cannot imply stronger RLDC lower bounds.

First, there exist t-daisies satisfying the condition $t|K| = O(n^{1-1/q})$ which are *not* p-robust daisies for any $p = n^{-o(1/q^2)}$. To see this, fix $t = n^{1-1/q}$, and consider the set system consisting of $n/t = n^{1/q}$ sets of size q, each repeating t times. This set system is a t-daisy with

[2] The robust daisy extracted in [GL21] is of density of only $\mu(\mathcal{D}) \geq 1/q$. Therefore, they had to employ a soundness amplification process, which increases the query complexity to $q \log q$. This increment gives the term $q^2 \log^2 q$ in their lower bound. Using an improved construction, [Gol23] is able to extract a robust daisy with density arbitrarily close to 1 (but still with $p = n^{-\Theta(1/q^2)}$), which avoids the need for soundness amplification, yielding the improved lower bound.

[3] Technically, the argument in [Gol23] avoids the notion of t-daisies, and the query sets are simply divided into heavy, medium, and light elements. However, the essence of the structure is maintained, as the heavy elements play the role of the kernel, and the light elements admit the bounded intersection property.

[4] Notice that extracting a t-daisy with $t|K| = O(n)$ is trivial, since for a constant q, there are at most $O(1/t)$ elements with degree larger than t.

an empty kernel, and satisfies $t|K| = O(n^{1-1/q})$. However, by a straightforward calculation, one needs $p = n^{-O(1/q^2)}$ to sample a full set with a constant probability.[5]

Nevertheless, one might wonder: is it possible to extract a t-daisy with a better relation between t and K? This might imply that this t-daisy is a p-robust daisy with $p = n^{-o(1/q^2)}$. Alas, this is impossible. The process for extracting t-daisies in [GL21] is optimal; one cannot find better t-daisies in arbitrary set systems.

To illustrate this, we present a $(q+1)$-uniform set system, with n sets over $O(n)$ base elements. This set system has the property that for every value of t, and setting the kernel K to contain the elements with degree larger than t, it holds that $t|K| = \Omega(n^{1-1/q})$. Consider the k-regular tree with $q+1$ levels, with $k = n^{1/q}$. Each level ℓ in the tree has $k^{\ell-1}$ vertices (the root is on level 1), and the last level where $\ell = q+1$ has n leaves. The n sets in the system are the unique n root-to-leaf paths in the tree. The degree of each element in level ℓ is $k^{q+1-\ell}$, which is the number of root-to-leaf paths going through it. The degrees decrease from $k^q = n$ at the root to $k^0 = 1$ at the leaves.

Now, consider any degree threshold t. This threshold must fall between the degrees of two adjacent levels. Namely, $t \in [k^{q-\ell}, k^{q+1-\ell})$ for some ℓ between 0 and q. The kernel K of a t-daisy now must include all vertices in the top ℓ levels – otherwise there is an element included in $k^{q+1-\ell} > t$ or more petals. The size of this kernel is therefore at least the total number of vertices in these levels:

$$|K| \geq \sum_{j=1}^{\ell} k^{j-1} = \Theta(k^{\ell-1}) \ .$$

On the other hand, $t \geq k^{q-\ell}$, and together this implies:

$$t|K| = \Omega(k^{q-\ell} \cdot k^{\ell-1}) = \Omega(k^{q-1}) = \Omega(n^{1-1/q}) \ .$$

To summarize: the notion of t-daisies is, on the one hand, not strong enough to imply the desired $n^{-\Theta(1/q)}$ sampling probability, and on the other hand, too strong, so improving the extraction process is impossible.

Nevertheless, as we shall see next, it is possible to extract a $n^{-\Theta(1/q)}$ robust daisy from every distribution supported over sets of size at most q, even from the tree set system above – this is guaranteed by our Robust Daisy Lemma (Lemma 1.2). For that, however, we need new ideas that avoid the bottleneck of t-daisies.

(m,k)-spreadness. In this work, we introduce the notion of (m,k)-*spreadness*, which is a generalization and strengthening of the known notion of k-spreadness [Tal10, MNWSZ25]. This stronger notion is essential for extracting robust daisies that outperform those from previous approaches.

In the following, we use $\langle T \rangle$ to denote the family of all subsets of U that contain T, and then $\mu(\langle T \rangle)$ is the total density of all subsets of U that contain T.

Definition 2.2 ((m,k)-spread distributions)**.** *Let μ be a distribution over $\mathcal{P}(U)$, let $k > 1$ and let $m \in (0,1]$. We say that μ is* (m,k)-spread *if for any non-empty set $T \subseteq U$,*

$$\mu(\langle T \rangle) \leq \frac{m}{k^{|T|}} \ .$$

[5] For a counterexample that avoids repetitions of the same set, one can add a unique vertex to each of the n sets, while maintaining the sampling bound.

This new notion coincides with the standard k-spreadness for $m = 1$. For $m < 1$, however, it is stronger. We believe this generalization, and the connection it creates between RLDCs to the literature on spreadness, might be of independent interest.

We remark that even with $m = 1$, the spreadness condition is strictly stronger than the condition required from the petals of a t-daisy. To see that, fix $m = 1$, and assume the uniform distribution over a set system $\mathcal{F}$ is k-spread. Then, applying the spreadness condition to $T = \{x\}$ for every universe element x, we deduce that every element is contained in at most $\frac{|\mathcal{F}|}{k}$ sets. That is, $\mathcal{F}$ is a $\frac{|\mathcal{F}|}{k}$-daisy with an empty kernel. In other words, the bounded intersection requirement is similar to asking for spreadness, but for singletons only. Spreadness is a much stronger requirement, as it applies to any subset of U.

Our generalized notion of (m, k)-spreadness allows us to prove a new spread lemma, which is useful specifically when the support of the distribution is over small sets. Before stating the new lemma, let us demonstrate why we need the new generalized definition.

(m, k)-spreadness vs k-spreadness. The main benefit of the new notion comes from the following observation: One cannot hope for spreadness parameter better than $k = n^{\Theta(1/q)}$. Consider the uniform distribution over n distinct sets, each of size q. Let T be one of these n sets. The probability of picking T is $\mu(T) = 1/n$. Since $T \in \langle T \rangle$, the total density $\mu(\langle T \rangle)$ must be at least $1/n$. On the other hand, the spreadness condition requires $\mu(\langle T \rangle) \leq 1/k^{|T|} = 1/k^q$. Combined, this implies $1/n \leq 1/k^q$, and hence $k \leq n^{1/q}$.

Now, the high-probability version of the spread lemma ([Rao20]) guarantees that a k-spread set system is (p, ε)-satisfying for $p = O\left(\frac{\log(q/\varepsilon)}{k}\right)$. This is known to have an optimal dependence on ε (e.g., see [BCW21, Lemma 4]). To obtain $\varepsilon = 2^{-\Omega(|K|)}$, one would need to set $p = n^{-\Theta(1/q)} \cdot |K|$ which is vacuous if $|K| = \omega(n^{1/q})$, which is unavoidable in certain cases. This motivates the need for a stronger notion of spreadness, and a stronger spread lemma.

A second difference between the definitions is that the new one is "closed under conditioning": if μ is (m, k)-spread, then, for any $\mathcal{D} \subseteq \mathrm{supp}(\mu)$, the conditioned distribution $\mu_{\mathcal{D}}$ is $\left(\frac{m}{\mu(\mathcal{D})}, k\right)$-spread. The more refined condition of (m, k)-spreadness allows us to express this relation. We will see shortly how this property helps us to show that a distribution is a robust daisy.

The Small-Set Spread Lemma. In Section 4, we prove that if a distribution is (m, k)-spread, then its support is a satisfying set system.

Lemma 2.3 ("The Small-Set Spread Lemma"; informally stated, see Lemma 4.6)**.** *Let μ be a (m, k)-spread distribution, and assume every set in $\mathrm{supp}(\mu)$ has at most q elements. Then, for every $\alpha > 2q$, $\mathrm{supp}(\mu)$ is (p, ε)-satisfying with $p = \frac{\alpha}{k}$ and $\varepsilon = \exp\left(-\widetilde{\Omega}\left(\frac{\alpha}{qm}\right)\right)$.*

This lemma shows that (m, k)-spreadness is sufficient to obtain the required satisfaction guarantees, and provides a clear target for what parameters we should aim for.

Suppose we can find a dense family $\mathcal{D}$ and a kernel K such that the distribution over the petals, μ, is (m, k)-spread with $k = n^{\Theta(1/q)}$ and $m|K| = O(1)$. Let $\mathcal{D}' \subseteq \mathcal{D}$. As noted above, $\mu_{\mathcal{D}'}$ is $(\frac{m}{\mu(\mathcal{D}')}, k)$-spread. We could then apply Lemma 2.3 with $\alpha = O(m|K|)$ (which is $O(1)$ by our assumption). The lemma then gives us sampling probability $p = O(1/k) = n^{-\Theta(1/q)}$ and failure probability $\varepsilon' = \exp\left(-\widetilde{\Omega}\left(\frac{m|K|}{qm/\mu(\mathcal{D}')}\right)\right) = 2^{-\Omega(\mu(\mathcal{D}') \cdot |K|)}$, implying that the set of

petals is $(n^{-\Theta(1/q)}, \varepsilon^{\mu(\mathcal{D}')})$-satisfying for $\varepsilon = 2^{-\Omega(|K|)}$, as needed. In the next section we show that we can always achieve such parameters.

Our proof of the Small-Set Spread Lemma relies on a delicate application of Janson's Inequality. We follow the proof strategy of prior spread lemmas [Ros14, ALWZ21], but crucially leverage our new definition of (m, k)-spread. We use this strengthened spread property to obtain a tighter bound on the cumulative dependency among intersecting sets in the family, which in turn yields a smaller failure probability.

While the above guarantee is stronger than that provided by standard k-spread, fortunately, the kernel structure of robust daises enable us to extract such a stronger spread in any distribution, as we discuss next.

2.3 The spreadness extraction lemma

By the discussion above, to obtain the desired RLDC lower bound, the remaining task is as follows. We get an arbitrary distribution μ over a universe U of size n, supported on sets of size at most q, which for the overview we assume is constant. We need to extract an (m, k)-spread distribution from μ, and we are allowed to perform the following two operations:

1. **Puncturing**: We remove a small set of "problematic" elements $K \subseteq U$, where $|K| = o(n)$.[6] This K corresponds to the kernel of the robust daisy.

2. **Conditioning**: We are allowed to restrict the distribution to a "well-behaved" subfamily $\mathcal{D} \subseteq \mathrm{supp}(\mu)$, as long as the density $\mu(\mathcal{D})$ is large (e.g., $\mu(\mathcal{D}) \geq 0.99$).

In Lemma 5.4, we prove that for any distribution μ, there *always exists* such a set K and a subfamily $\mathcal{D}$ which allow us to achieve the desired spreadness.

Lemma 2.4. *(The Spreadness Extraction Lemma; informally stated, see Lemma 5.4) Let μ be a distribution over $\mathcal{P}(U)$, and assume that every set in $\mathrm{supp}(\mu)$ has at most q elements.*

Then, there exists a family of sets $\mathcal{D} \subseteq \mathrm{supp}(\mu)$, and a set $K \subseteq U$ such that the distribution $\mu_{\mathcal{D}}$ punctured by K is (m, k)-spread, where:

$$k = n^{\Theta(1/q)}, \qquad m \cdot |K| = O(1), \qquad \mu(\mathcal{D}) \geq 0.99 \ .$$

This extraction process builds upon [GL21] and [Gol23]. However, it is substantially more involved, as we extract (m, k)-spreadness, as opposed to merely a degree bound.

We next provide a high-level overview of the proof of Lemma 2.4.

Universe partitioning. We partition the elements in U into $c + 1 = \Theta(q)$ buckets. The partitioning is according to the (normalized) *weighted degree* of each element, which we define as:

$$\bar{\mu}(u) = \sum_{S \text{ s.t. } u \in S} \frac{\mu(S)}{|S|} \ .$$

Note that $\bar{\mu}$ is a distribution over U. That is, $\sum_{x \in U} \bar{\mu}(x) = 1$. This distribution is equivalent to the following two-step random process: first, select a set $S \subseteq U$ according to μ, and then select an element $x \in S$ uniformly at random. We remark that if μ is a uniform distribution

[6] We note that it is sufficient for our purposes that $|K| \leq \delta n$. However, for a constant q, we can achieve this better $o(n)$ bound.

over a uniform family $\mathcal{F}$, then $\bar{\mu}(u)$ is the normalized degree of u in $\mathcal{F}$ (the fraction of sets in $\mathcal{F}$ containing u) divided by q.

We set a *base step size* $k = n^{1/c}$. The buckets are then defined by degree ranges: B_0 contains the elements with $\bar{\mu}(u) \leq 1/n$, and for $j \in [c]$, the bucket B_j contains the elements with $\bar{\mu}(u) \in (k^{j-1}/n, k^j/n]$. Since $k^c = n$ and the degrees are normalized, this is indeed a partition of the universe.

The kernel threshold. We choose the kernel K to be all elements with $\bar{\mu}(u) > k^j/n$ for some $j \in [c]$ that we pick later. That is, the *threshold* $m = k^j/n$ is set to be one of the bucket boundaries. Since $\sum_{x \in U} \bar{\mu}(x) = 1$, this implies $m|K| = O(1)$. The question is how to choose this bucket j (which defines K and m), such that after puncturing by K and conditioning on an appropriate subfamily $\mathcal{D}$ (as we will see next), the resulting distribution is $(O(m), k)$-spread.

Token distributions and good boundaries. For any fixed set S, the partitioning categorizes its q elements into the buckets. We treat these elements as q *tokens* distributed among the $c+1$ buckets, and call this distribution the *token distribution* of S.

Now, we define a key property: we say that a bucket B_j is a *good boundary* for S if the token distribution satisfies the following property: for every $i \in \{0, \ldots, j\}$, the set of $i+1$ buckets from B_{j-i} to B_j (inclusive) contains at most i tokens. The proof relies on the following two claims:

1. If B_j is a good boundary for every $S \in \mathcal{D}$ (for some $\mathcal{D} \subseteq \mathrm{supp}(\mu)$), then after the removal of $K = \{u \mid \bar{\mu}(u) > k^j/n\}$ the distribution $\mu_{\mathcal{D}}$ punctured by K is $(O(m), k)$-spread.
2. There exists $\mathcal{D} \subseteq \mathrm{supp}(\mu)$ with $\mu(\mathcal{D}) > 0.99$ and a bucket B_j which is a good boundary for every $S \in \mathcal{D}$.

We next sketch the proofs of these two claims, which form the technical heart of the lemma.

From good boundaries to spreadness. Recall that to prove spreadness, we need to show that for every $T \subseteq U$, the total mass of sets containing T (denoted by $\mu(\langle T \rangle)$) is upper bounded by $m/k^{|T|}$. The key observation is that: if T *contains an element with small weighted degree*, then $\mu(\langle T \rangle)$ itself is upper bounded. Specifically, let $x \in T$. Then, since each set containing T also contains x, and each set is of size at most q, we have $\mu(\langle T \rangle) \leq q \cdot \bar{\mu}(x)$.

Now, since we removed all elements with $\bar{\mu}(u) > k^j/n$, we can assume T does not contain any such element. But, by construction, B_j is a good boundary, hence, by taking $i = |T| - 1$, T can contain at most $|T| - 1$ elements in the buckets between $B_{j-|T|+1}$ and B_j (inclusive). In other words, T must contain an element in $B_{j-|T|}$ (or a lower bucket). But all these elements have $\bar{\mu}(u) \leq k^{j-|T|}/n = m/k^{|T|}$. Hence, by the above argument, we get the bound $\mu(\langle T \rangle) \leq qm/k^{|T|}$.

Abundance of good buckets. We prove that for a any fixed set S, at least $c - q$ buckets constitute good boundaries. Towards this end, we use a subtle analysis of a *token shifting* process: by iteratively moving tokens from crowded buckets to higher ones, we can argue that the final configuration will have at most q buckets containing any tokens. This leaves at least $c - q$ buckets empty, and we prove that these empty buckets must be good boundaries.

Taking a sufficiently large c (e.g., $c = 100q = \Theta(q)$) yields that almost all the boundaries (say, 99%) are good. In turn, this implies that there exists (at least) one bucket which is a good boundary for 99% of the sets. We choose this bucket to determine m and K, and set $\mathcal{D}$ to be the sets for which this bucket is good.

The number of buckets. It is instructive to discuss how we choose the exact number of buckets, c. On the one hand, we want this number to be as small as possible, as c determines the exact constant in the exponent of the lower bound. Specifically, the lower bound we achieve is $n = \widetilde{\Omega}(k^{1+\frac{1}{c-1}})$.

On the other hand, it must be sufficiently large; the fraction of good boundaries (out of the total number of buckets) must be larger than the soundness error of the relaxed decoder, which is $1-\sigma$. Otherwise, even if the chosen boundary is good for a large fraction of sets, it is possible that all these sets are "bad" — leading the decoder to a wrong output when choosing them. This means we must choose c such that the fraction of good boundaries $\frac{c-q}{c} = 1 - q/c$ is larger than $1-\sigma$; that is, $q/c < \sigma$. More precisely, we need σ to be bounded away from q/c by some constant independent of k.

Hence, we choose $c = \lceil \frac{q}{\sigma} \rceil + 1$, which is the minimal integer satisfying this requirement. This parameter choice gives the exact exponent in the lower bound – namely, $\frac{1}{\lceil q/\sigma \rceil}$.

The centrality of the spread parameter $k = n^{1/q}$. The choice for the value of k, the "spreadness" parameter or "step size", plays a central role in determining the exact lower bound achieved, even more than the choice of c. Let us follow how this parameter propagates through the different parts of the proof.

First, the Small-Set Spread Lemma yields a (p, ε)-satisfying set system with $p = \alpha/k$. Recall we set $\alpha = O(1)$ in this overview, so $p = O(1/k)$.[7] This sampling probability p is then used for the robust daisy we extract. By our reduction (Lemma 2.1), this value of p is what gives the final lower bound on the code's length.

Second, we remark that our current proof technique cannot get an improved value for k. As we have seen, $k = n^{1/c}$ is the base step size between the buckets, and our proof requires $c > q$. If there were fewer than $q+1$ buckets (i.e., $c \leq q$), it would be possible for a set S to place one of its q tokens in every single bucket. In this case, our token-shifting argument would fail to guarantee an empty bucket, and there would be no "good boundaries" at all.

This $k = n^{\Theta(1/q)}$ barrier is not a coincidence; it appears to be fundamental. This exponent is not just a limitation of our proof technique but an inherent feature of the problem, stemming from three different points of view:

1. **From Construction:** It matches the $n^{1+O(1/q)}$ upper bound, so a better parameter would imply an impossible lower bound.

2. **From Combinatorics:** As observed above, $k = n^{\Theta(1/q)}$ is the best spreadness parameter achieved for arbitrary distributions over sets of size q.

3. **From Our Proof:** Our token-shifting technique, which requires $c > q$ buckets, independently arrives at the same $k = n^{\Theta(1/q)}$ parameter.

[7] More accurately, in the full proof we need to set $\alpha = O(\log^2 n \log^2 |\Sigma|)$ to compensate for a few minor terms which we do not cover in the overview.

These perspectives – the upper bound, the combinatorial limitation of spreadness, and our own extraction method – solidify the $1/q$ exponent as a central, inherent property of the problem.

2.4 Organization

The rest of the paper is organized as follows. In Section 3 we give standard definitions and notations, including those of RLDCs. In Section 4 we prove the spread lemmas, making the link between spreadness and robust daisies. In Section 5 we prove the spreadness extraction lemma, and apply it to prove the robust daisy lemma. In Section 6 we show the reduction from proving a lower bound for RLDCs to the problem of finding a robust daisy, and we apply it to finally prove Theorem 1.

3 Preliminaries

We provide notation for set and distributions that we shall use throughout the paper.

3.1 Basic notations

Set notation. Let U be a finite set.

- For a set U, we denote by $\mathcal{P}(U)$ the power set of U, i.e., the family of all subsets of U. Furthermore, let $\mathcal{P}_{\leq q}(U) = \{S \in \mathcal{P}(U) \ ; \ |S| \leq q\}$ denote the family of subsets of size at most q.
- The *star* of a set $T \subseteq U$, denoted by $\langle T \rangle$, is the family of subsets of U that contain T; that is, $\langle T \rangle = \{S \subseteq U \mid S \supseteq T\} \subseteq \mathcal{P}(U)$.

  We use a slight abuse of notation and write $\langle x \rangle$ to denote $\langle \{x\} \rangle$ for an element $x \in U$.
- A family of sets $\mathcal{F} \subseteq \mathcal{P}(U)$ is called *q-uniform* if every set $S \in \mathcal{F}$ has size $|S| = q$.
- We use $\mathbb{I}[\cdot]$ to denote the *indicator function*. For a set S, the indicator function of S is defined as:

$$\mathbb{I}[x \in S] = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

Distributions. Let D be a discrete domain.

- A *distribution* μ over D is a function $\mu : D \to [0,1]$ such that $\sum_{x \in D} \mu(x) = 1$.

  For any subset $A \subseteq D$, the *probability mass* or *density* of A is $\mu(A) = \sum_{x \in A} \mu(x)$.
- The *support* of a distribution μ is the set of elements with non-zero probability, denoted $\mathrm{supp}(\mu) = \{x \in D \mid \mu(x) > 0\}$.
- The *conditioning* of a distribution μ to a set $A \subseteq D$ with non-zero density is a distribution over A, denoted μ_A, defined for each $x \in A$ by $\mu_A(x) = \frac{\mu(x)}{\mu(A)}$. The following straightforward observation relates the probability of events in the conditional distribution to the original distribution.

  Fact 3.1. *For any $B \subseteq D$, $\mu_A(B) = \mu(B)/\mu(A)$.*

3.2 Concentration inequalities

We use the following version of Janson's inequality.

Lemma 3.2 (Janson's Inequality [AS16, Chapter 8.1])**.** *Let $\mathcal{S} \subseteq \mathcal{P}(U)$ be a family of sets. Let $W \sim \mathrm{Bin}(U,p)$. For $S \in \mathcal{S}$, let Z_S be the indicator of the event that $S \subseteq W$. Let $X = \sum_{S\in\mathcal{S}} Z_S$ and $M = \mathbb{E}[X]$. For, $S \neq T \in \mathcal{S}$, let $S \sim T$ iff $S \cap T \neq \emptyset$. Define*

$$\Delta = \sum_{(S,T)\,:\,S\sim T} \mathbb{E}[Z_S Z_T] \,.$$

Then,

$$\Pr[X = 0] \leq \exp(-M + \Delta/2) \,,$$

and if $M \leq \Delta$, then

$$\Pr[X = 0] \leq \exp(-M^2/2\Delta) \,.$$

3.3 Relaxed Locally Decodable Codes

We recall the formal definition of relaxed locally decodable codes [BGH+06].

Definition 3.3. *(Relaxed locally decodeable codes) Let $C : \{0,1\}^k \to \Sigma^n$ be an error correcting code. A (q,δ,σ)-relaxed decoder for C is a randomized procedure $\mathcal{B}$ that on an explicit input $i \in [k]$, and oracle access to $w \in \Sigma^n$, outputs an element of $\{0,1,\perp\}$, and satisfies the following requirements:*

1. *(completeness) If $w = C(x)$ for some $x \in \{0,1\}^k$, then $\mathcal{B}^w(i) = x_i$.*[8]
2. *(relaxed local decoding) If there exists $x \in \{0,1\}^k$ such that $\mathrm{dist}(w, C(x)) \leq \delta$, then $\mathcal{B}^w(i) \in \{x_i, \perp\}$ with probability at least σ.*
3. *For every input i and oracle access to any $w \in \Sigma^n$, $\mathcal{B}$ makes at most q queries.*

We say that a $\mathcal{B}$ is non-adaptive if it determines all its queries based on its explicit input (namely, the index to decode) and internal coin tosses, independently of the specific w to which it is given oracle access. We refer to δ as the decoding radius of the decoder, and to σ as its soundness probability.

A code C with a (q,δ,σ)-relaxed decoder is often referred to as a (q,δ,σ)-relaxed locally decodable code.

A relaxed *corrector* for a code is defined analogously to a relaxed decoder, but its objective is to correct any codeword symbol rather than decode a message bit. That is, the three requirements in Definition 3.3 are extended to any $i \in [n]$, and the corrector's output should be c_i (or $\perp$) instead of x_i.

We say that a code $C : \{0,1\}^k \to \Sigma^n$ is *linear* if Σ is a field and the image of C is a linear subspace of Σ^n.

[8] We remark that a common variation of RLDCs allows the decoder to err with a small probability even on valid codewords. Our lower bound holds for such codes as well (assuming they are linear), but for simplicity, we assume perfect completeness throughout.

4 Spread lemmas

In this section, we prove two spread lemmas. The first is a spread lemma for *families of sets*. We then proceed to extend this lemma, and prove a spread lemma for distributions.

Let us begin by reiterating the relevant definitions.

Definition 4.1 (Satisfying set system)**.** *Let $\mathcal{F}$ be a family of sets over a universe U. We say that $\mathcal{F}$ is (p,ε)-satisfying if*

$$\Pr_{W\sim Bin(U,p)}[\exists S\in\mathcal{F}, S\subseteq W]\geq 1-\varepsilon\,.$$

Here, $W\sim Bin(U,p)$ denotes the random set where each element $u\in U$ is chosen to be in W independently with probability p.

We remark that the notion of a satisfying set system, and its name, originates from the study of DNF formulas. When a set system is interpreted as a DNF formula, this condition is that the formula has more than a $1-\varepsilon$ probability of being satisfied on p-biased inputs.

In the following, recall that for $T\subseteq U$, the star of T, denoted $\langle T\rangle$, is the family of all subsets of U that contain T.

Definition 4.2 ((m,k)-spread distributions, reiterating Definition 2.2)**.** *Let μ be a distribution over $\mathcal{P}(U)$, let $k>1$ and let $m\in(0,1]$. We say that μ is* (m,k)-spread *if for any non-empty set $T\subseteq U$,*

$$\mu(\langle T\rangle)\leq\frac{m}{k^{|T|}}\,.$$

Definition 4.3 (Spread families)**.** *Let $\mathcal{F}$ be a family of sets over a universe U, let $k>1$, and let $m\in(0,1]$. We say that $\mathcal{F}$ is (m,k)-spread if the uniform distribution over $\mathcal{F}$ is (m,k)-spread. That is, if for any non-empty set $T\subseteq U$,*

$$\frac{\deg_{\mathcal{F}}(T)}{|\mathcal{F}|}\leq\frac{m}{k^{|T|}}\,,$$

where $\deg_{\mathcal{F}}(T)$ is the number of sets in $\mathcal{F}$ that contain T.

4.1 The spread lemma for families of sets

We start by proving the spread lemma for fixed set systems.

Lemma 4.4 ("The Small-Set Spread Lemma" - for families of sets)**.** *Fix $k>1$ and $m\in(0,1]$. Let $\mathcal{F}$ be an (m,k)-spread family of sets over universe U, and assume every set in $\mathcal{F}$ has at most q elements. Then, for every $\alpha>2q$, the family $\mathcal{F}$ is (p,ε)-satisfying with $p=\alpha/k$ and $\varepsilon=\exp\left(-\frac{\alpha}{4qm}\right)$.*

Proof. Our goal is to lower bound $\Pr_{W\sim\mathrm{Bin}(U,p)}[\exists S\in\mathcal{F}, S\subseteq W]$, where $p=\alpha/k$. The proof relies on Janson's inequality.

Setup of Janson's inequality. For any $S \in \mathcal{F}$, let $\mathcal{Z}_S$ be the indicator value for the event $S \subseteq W$. Denote $S \sim T$ if $S, T \in \mathcal{F}$ intersect. Define:

$$M = \sum_{S \in \mathcal{F}} \mathbb{E}[\mathcal{Z}_S], \quad \Delta = \sum_{S \sim T} \mathbb{E}[\mathcal{Z}_S \mathcal{Z}_T] \,. \tag{1}$$

Recall that Janson's inequality states that:

1. when $\Delta \leq M$:

$$\Pr_{W \sim \mathrm{Bin}(U,p)}[\exists S \in \mathcal{F}, S \subseteq W] \geq 1 - \exp\left(-M + \Delta/2\right) \geq 1 - \exp(-M/2).$$

2. when $\Delta > M$:

$$\Pr_{W \sim \mathrm{Bin}(U,p)}[\exists S \in \mathcal{F}, S \subseteq W] \geq 1 - \exp\left(-\frac{M^2}{2\Delta}\right).$$

Let $s \leq q$ be the size of the largest set in $\mathcal{F}$. It is sufficient to show that

$$\min\left(\frac{M^2}{2\Delta}, \frac{M}{2}\right) \geq \frac{\alpha}{4sm} \geq \frac{\alpha}{4qm} \,.$$

Uniformity assumption. In what follows, we assume that $\mathcal{F}$ is s-uniform. This assumption can be made without loss of generality. If the family is not uniform, we can increase the universe with a set of "dummy" elements and pad each set $S \in \mathcal{F}$ with $s - |S|$ distinct dummies to create a new s-uniform family $\mathcal{F}'$. This transformation only makes the required condition stricter: for any sample W from the new universe, if a padded set $S' \in \mathcal{F}'$ is fully contained in W, then its original counterpart S is necessarily contained in W as well.

Also note that since each new dummy element is contained only in a single set, this transformation does not affect the spreadness of $\mathcal{F}$.

Estimating M. By assumption, every set in $\mathcal{F}$ has s elements, and hence for any $S \in \mathcal{F}$ we have $\mathbb{E}[\mathcal{Z}_S] = p^s$. Therefore,

$$M = \sum_{S \in \mathcal{F}} \mathbb{E}[\mathcal{Z}_S] = |\mathcal{F}| \cdot p^s \,. \tag{2}$$

Bounding Δ. For any $t \in [s]$, let r_t be the number of pairs of sets $S, T \in \mathcal{F}$ such that $|S \cap T| = t$. Any such $S, T \in \mathcal{F}$ share t vertices, and have $s - t$ unique vertices each. Therefore, the probability of sampling all of the vertices of both S, T is

$$\mathbb{E}[\mathcal{Z}_S \mathcal{Z}_T] = p^t \cdot (p^{s-t})^2 = p^{2s-t} \,.$$

Hence, it follows that

$$\Delta = \sum_{S \sim T} \mathbb{E}[\mathcal{Z}_S \mathcal{Z}_T] = \sum_{t=1}^{s} r_t \cdot p^{2s-t} = p^{2s} \sum_{t=1}^{s} r_t \cdot p^{-t} \,.$$

We proceed to bound r_t. For each of the $|\mathcal{F}|$ subsets in the family, there are $\binom{s}{t} \leq s^t$ options to choose the intersection set $R \subseteq S$ of size t. By the spreadness hypothesis, each

such set R is contained in at most $\frac{m}{k^t}|\mathcal{F}|$ sets of $\mathcal{F}$. Therefore, in total, $r_t \leq |\mathcal{F}|^2 m \left(\frac{s}{k}\right)^t$, and we get the upper bound

$$\Delta \leq p^{2s} \sum_{t=1}^{s} |\mathcal{F}|^2 m \left(\frac{s}{k}\right)^t \cdot p^{-t} = mp^{2s}|\mathcal{F}|^2 \sum_{t=1}^{s} \left(\frac{s}{pk}\right)^t .$$

Next, since $pk = \alpha$ and by assumption $\alpha > 2s \implies (1 - s/\alpha) > 1/2$:

$$\Delta \leq mp^{2s}|\mathcal{F}|^2 \sum_{t=1}^{s} \left(\frac{s}{pk}\right)^t \leq mp^{2s}|\mathcal{F}|^2 \sum_{t=1}^{\infty} \left(\frac{s}{\alpha}\right)^t = mp^{2s}|\mathcal{F}|^2 \left(\frac{s/\alpha}{1 - s/\alpha}\right) \leq mp^{2s}|\mathcal{F}|^2 \frac{2s}{\alpha} . \quad (3)$$

Applying Janson's inequality. Combining Equation (2) and Equation (3), we conclude that

$$\frac{M^2}{2\Delta} \geq \frac{(|\mathcal{F}|p^s)^2}{2mp^{2s}|\mathcal{F}|^2\alpha^{-1}2s} = \frac{\alpha}{4sm} .$$

Also, note that

$$\frac{M}{2} = \frac{p^s|\mathcal{F}|}{2} = \frac{\alpha^s|\mathcal{F}|}{2k^s} \geq \frac{\alpha}{2} \cdot \frac{|\mathcal{F}|}{k^s}$$

where the last inequality follows since $\alpha \geq 1$. Now, by assumption, there exists a set $S \in \mathcal{F}$ (without dummy elements) with $|S| = s$. Hence, from the spreadness of $\mathcal{F}$ when applied to $T = S$, we get $1 \leq \deg(S) \leq |\mathcal{F}| \cdot \frac{m}{k^s} \implies \frac{|\mathcal{F}|}{k^s} \geq \frac{1}{m}$, and then:

$$\frac{M}{2} \geq \frac{\alpha}{2} \cdot \frac{1}{m} \geq \frac{\alpha}{4sm}$$

□

4.2 The spread lemma for distributions

We use the following helper lemma to transition from distributions to fixed set systems.

Lemma 4.5. *Let μ be a distribution over a support R with $|R| \geq 2$. There exists a non-empty set $A \subseteq R$ such that for every $a \in A$:*

$$\mu(a) \geq \frac{1}{2|A|\log(|R|)}$$

Proof. Let $n = |R|$, and let $p_1 \geq p_2 \geq \cdots \geq p_n$ be the sorted probabilities of the elements in R. Assume for contradiction that for all $k \in \{1, \ldots, n\}$, we have $p_k < \frac{1}{2k\log(n)}$.

Summing over all k gives:

$$1 = \sum_{k=1}^{n} p_k < \sum_{k=1}^{n} \frac{1}{2k\log(n)} = \frac{1}{2\log(n)} \sum_{k=1}^{n} \frac{1}{k} = \frac{H_n}{2\log(n)}$$

This implies $2\log(n) < H_n$. However, for $n \geq 2$, it is a known that $H_n \leq \ln(n) + 1 \leq 2\log(n)$, which is a contradiction.

Therefore, there must exist an index k such that $p_k \geq \frac{1}{2k\log(n)}$. Let A be the set of the k elements with the largest probabilities. Then $|A| = k$, and for any $a \in A$, its probability $\mu(a) \geq p_k$, which satisfies the desired bound. □

Now, we prove the full version of our small-set spread lemma.

Lemma 4.6 ("The Small-Set Spread Lemma")**.** *Fix $k > 1$ and $m \in (0, 1]$. Let μ be a (m, k)-spread distribution over $\mathcal{P}(U)$ with support $\mathcal{F}$, and assume every set in $\mathcal{F}$ has at most q elements. Then, for every $\alpha > 2q$, $\mathcal{F}$ is (p, ε)-satisfying with $p = \alpha/k$ and $\varepsilon = \exp\left(-\frac{\alpha}{8qm \log|\mathcal{F}|}\right)$.*

Proof. First, observe that if $\mathcal{F}' \subseteq \mathcal{F}$ is (p, ε)-satisfying, then $\mathcal{F}$ is also (p, ε)-satisfying.

Now, by Lemma 4.5, there is a subset $\mathcal{F}' \subseteq \mathcal{F}$ such that for every $S \in \mathcal{F}'$:

$$\mu(S) \geq \frac{1}{2|\mathcal{F}'| \log|\mathcal{F}|} \implies 1 \leq 2|\mathcal{F}'| \log|\mathcal{F}| \mu(S)$$

We argue that $\mathcal{F}'$ is $(2\log|\mathcal{F}|m, k)$-spread. Let $T \subseteq U$ be a non-empty set. Then,

$$\begin{aligned}\frac{\deg_{\mathcal{F}'} T}{|\mathcal{F}'|} &= \frac{\sum_{S \in \langle T \rangle} \mathbb{I}[S \in \mathcal{F}']}{|\mathcal{F}'|} \\ &\leq \frac{\sum_{S \in \langle T \rangle} 2|\mathcal{F}'| \log|\mathcal{F}| \mu(S)}{|\mathcal{F}'|} \\ &= 2\log|\mathcal{F}| \mu(\langle T \rangle) \leq 2\log|\mathcal{F}| \cdot m \cdot k^{-|T|}\end{aligned}$$

where in the last inequality we applied the spreadness of μ.

Now, applying Lemma 4.4, and as every set in $\mathcal{F}'$ has at most q elements, we conclude that $\mathcal{F}'$ is (p, ε)-satisfying (and hence also $\mathcal{F}$) for $p = \alpha/k$ and

$$\varepsilon = \exp\left(-\frac{\alpha}{4q \cdot 2\log|\mathcal{F}|m}\right) = \exp\left(-\frac{\alpha}{8qm \log|\mathcal{F}|}\right)$$

□

5 Robust daisies

In this section, we prove the Robust Daisy Lemma: every distribution over sets of small size contains a dense robust daisy (Definition 1.1).

Lemma 5.1. *(The Robust Daisy Lemma) Let μ be a distribution over $\mathcal{P}_{\leq q}(U)$. For every integer $c > q$ there exists $\mathcal{D} \subseteq \mathcal{P}_{\leq q}(U)$ such that for every $\alpha > 2q$, the conditioned distribution $\mu_{\mathcal{D}}$ is a (p, ε)-robust daisy with a non-empty kernel $K \subseteq U$ where*

$$p = \frac{\alpha}{n^{1/c}} \qquad \varepsilon = \exp\left(-\frac{\alpha(1 - q/c)}{8q^2 \log|\mathcal{D}|} \cdot |K|\right) \qquad |K| \leq n^{1-1/c} \qquad \mu(\mathcal{D}) \geq 1 - \frac{q}{c}.$$

The proof of the lemma is in two steps. First, we prove the "Spreadness Extraction Lemma", Lemma 5.4. We show that it is always possible to make a distribution into a spread one by removing a kernel ("puncturing") and conditioning on a dense subfamily of the original support. This lemma is the main technical part of this section, and is proved in Section 5.1.

The second part of the proof is to apply the small-set spread lemma (Lemma 4.6) to argue that the extracted spread structure is a robust daisy with the required parameters.

5.1 The spreadness extraction lemma

We start by formally defining the puncturing operator, and observe its properties.

Definition 5.2 (Punctured Distribution)**.** *Let μ be a distribution over $\mathcal{P}(U)$. The* punctured distribution *of μ with respect to a set $K \subseteq U$ is a distribution over $\mathcal{P}(U \setminus K)$, denoted $\mu^{\circ K}$. It is defined by the process of first selecting a set $S \subseteq U$ according to μ and then outputting the set $S \setminus K$. The probability of a set $A \subseteq U \setminus K$ is given by*

$$\mu^{\circ K}(A) = \sum_{B \subseteq K} \mu(A \cup B).$$

Note that $\sum_{A \in \mathcal{P}(U \setminus K)} \mu^{\circ K}(A) = 1$, and so $\mu^{\circ K}$ is indeed a distribution over $\mathcal{P}(U \setminus K)$.

The following observations about the behavior of puncturing will be useful.

Observation 5.3. *For every distribution μ over $\mathcal{P}(U)$, a family $\mathcal{D} \subseteq \mathcal{P}(U)$ and a set $K \subseteq U$, the following holds:*

1. *If $\mathrm{supp}(\mu) = \mathcal{D}$ then $\mathrm{supp}(\mu^{\circ K}) = \mathcal{D} \setminus K = \{S \setminus K \mid S \in \mathcal{D}\}$.*
2. *If μ is supported on sets of size at most q, then so is $\mu^{\circ K}$.*
3. *$\mu^{\circ K}(\mathcal{D} \setminus K) \geq \mu(\mathcal{D})$.*
4. *For every $T \subseteq U \setminus K$, it holds that $\mu^{\circ K}(\langle T \rangle) = \mu(\langle T \rangle)$.*

Proof. Items 1-3 follow directly from the definitions. Item 4 follows by carefully expanding the definitions:

$$\mu^{\circ K}(\langle T \rangle) = \sum_{A \text{ s. t } T \subseteq A \subseteq U \setminus K} \mu^{\circ K}(A) = \sum_{A \text{ s. t } T \subseteq A \subseteq U \setminus K} \sum_{B \subseteq K} \mu(A \cup B)$$

Now observe that the double summation is the same as summing over all subsets of U that contain T. □

We are now ready to state and prove the spreadness extraction lemma.

Lemma 5.4. *(The Spreadness Extraction Lemma) Let μ be a distribution over $\mathcal{P}_{\leq q}(U)$, where $|U| = n$. Let $c > q$ be any integer.*

Then, there exists $j \in [c]$, a family of sets $\mathcal{D} \subseteq \mathrm{supp}(\mu)$, and a non-empty set of elements $K \subseteq U$ such that the distribution $(\mu_{\mathcal{D}})^{\circ K}$ is $(m, n^{1/c})$-spread, where:

$$m = \frac{q}{\mu(\mathcal{D})} \cdot \frac{n^{j/c}}{n}, \qquad |K| \leq \frac{n}{n^{j/c}}, \qquad \mu(\mathcal{D}) \geq 1 - \frac{q}{c}\,.$$

We note that when q and c are treated as constants, the parameters satisfy the relation $m \cdot |K| = O(1)$, which is needed for the robust daisy extraction. The specific value of the index j is not crucial beyond its role in defining m and K.

Proof. The proof proceeds by partitioning the universe U into $c+1$ buckets based on the *weighted degree* of each element, $\bar{\mu}$. We define $\bar{\mu}$ for each $x \in U$ by:[9]

$$\bar{\mu}(x) = \sum_{S \in \langle x \rangle} \frac{\mu(S)}{|S|} \ .$$

Note that $\bar{\mu}$ is a distribution over U. That is, $\sum_{x \in U} \bar{\mu}(x) = 1$. This distribution is equivalent to the following two-step random process: first, select a set $S \subseteq U$ according to μ, and then select an element $x \in S$ uniformly at random.

The construction identifies a threshold defined by one of the buckets, indexed by j. The punctured set K consists of all elements with a weighted degree above this threshold.

Construction. Let $k = n^{1/c}$. We partition the universe U into $c+1$ buckets, $\{B_0, B_1, \ldots, B_c\}$, based on the weighted degree $\bar{\mu}(u)$. Define $B_0 = \left\{u \in U \mid \bar{\mu}(u) \leq \frac{1}{n}\right\}$ and for each $j \in [c]$:

$$B_j = \left\{ u \in U \mid \bar{\mu}(u) \in \left(\frac{k^{j-1}}{n}, \frac{k^j}{n} \right] \right\} \ . \tag{4}$$

Since $k^c = n$, the upper bound for $\bar{\mu}(u)$ in B_c is 1. As $\bar{\mu}(u) \leq 1$ for any $u \in U$, the buckets $\{B_0, B_1, \ldots, B_c\}$ form a partition of U.

For any indices $i \leq \ell$, we denote $B_{[i,\ell]} = \cup_{j=i}^{\ell} B_j$.

The core of our argument relies on identifying a "well-behaved" boundary between buckets. We formalize this in the following definition.

Definition 5.5. *(Good Boundary) Let $j \in [c]$ and $S \subseteq U$. We say that the bucket B_j is a* good boundary *for the set S if for every $i \in \{0, \ldots, j\}$, we have $|S \cap B_{[j-i,j]}| \leq i$.*[10]

Note that by definition, B_0 cannot be a good boundary.

The lemma now follows from the three claims below, and setting $K = B_{[j+1,c]}$. It is possible that $B_{[j+1,c]}$ is empty, in which case we set $K = \{u\}$ for an arbitrary $u \in U$. This does not affect any of our arguments.

Claim 5.6. *There exists an index $j \in [c]$ and a family $\mathcal{D} \subseteq \mathcal{P}(U)$ such that B_j is a good boundary for every $S \in \mathcal{D}$, and $\mu(\mathcal{D}) \geq 1 - \frac{q}{c}$.*

Claim 5.7. *If B_j is a good boundary for every $S \in \mathcal{D}$, then the distribution $(\mu_{\mathcal{D}})^{\circ B_{[j+1,c]}}$ is $\left(\frac{q}{\mu(\mathcal{D})} \cdot \frac{k^j}{n} \ , \ k\right)$-spread.*

Claim 5.8. *For every index $j \in [c]$, the set of elements with high weighted degree is small: $|B_{[j+1,c]}| \leq \frac{n}{k^j}$.*

We begin by proving the claim that motivates the definition of good boundaries: if a bucket is a good boundary for some sets $\mathcal{D}$, then by conditioning on $\mathcal{D}$, and then removing the elements "above" this bucket, we get a spread distribution.

Proof of Claim 5.7. Assume that B_j is a good boundary for every $S \in \mathcal{D}$. We prove that the distribution $(\mu_{\mathcal{D}})^{\circ B_{[j+1,c]}}$ is $\left(\frac{q}{\mu(\mathcal{D})} \cdot \frac{k^j}{n} \ , \ k\right)$-spread.

[9] Recall that $S \in \langle x \rangle \iff x \in S$.

[10] Note that for $i \geq q$, this condition becomes trivial, as $|S| \leq q$ for every $S \in \mathrm{supp}(\mu)$.

Restating the goal. The distribution $(\mu_{\mathcal{D}})^{\circ B_{[j+1,c]}}$ is over subsets of

$$U \setminus B_{[j+1,c]} = B_{[0,j]} = \left\{ u \in U \mid \bar{\mu}(u) \leq \frac{k^j}{n} \right\} .$$

Hence, to show it is spread, we need to show that for every non-empty set $T \subseteq B_{[0,j]}$ we have

$$(\mu_{\mathcal{D}})^{\circ B_{[j+1,c]}}(\langle T \rangle) \leq \frac{q}{\mu(\mathcal{D})} \cdot \frac{k^j}{n} \cdot k^{-|T|}$$

By Observation 5.3 (item 4) and the definition of $\mu_{\mathcal{D}}$, this is equivalent to:

$$\mu(\langle T \rangle) \leq q \cdot \frac{k^{j-|T|}}{n} .$$

To show this, we first bound $\bar{\mu}(x)$ for some $x \in T$. Then, we use this bound to bound $\mu(\langle T \rangle)$.

Bounding $\bar{\mu}(x)$ for $x \in T$. We may assume there exists a set $S \in \mathcal{D}$ with $T \subseteq S$; otherwise, $\mu_{\mathcal{D}}(\langle T \rangle) = 0$ and the spreadness requirement holds trivially.

First, we show that $|T| \leq j$. Since B_j is a good boundary, $|S \cap B_{[j-i,j]}| \leq i$ for every $i \leq j$. Specifically, setting $i = j$ we have $|S \cap B_{[0,j]}| \leq j$. Now, because S is a subset of $B_{[0,j]}$, this simplifies to $|S| \leq j$, and therefore $|T| \leq j$.

Next, we apply the same inequality with $i = |T| - 1$. Note that $|T| \leq j$ ensures $i \leq j$. This yields:

$$\left| T \cap B_{[j-|T|+1,j]} \right| \leq |T| - 1 .$$

The inequality implies that at most $|T|-1$ elements of T belong to the set $B_{[j-|T|+1,j]}$. By the pigeonhole principle, at least one element $x \in T$ must lie outside this set. Since all elements of T are in $B_{[0,j]}$, x must be in $B_{[0,j]} \setminus B_{[j-|T|+1,j]} = B_{[0,j-|T|]}$. By definition, this means $\bar{\mu}(x) \leq \frac{k^{j-|T|}}{n}$.

Bounding $\mu(\langle T \rangle)$. First, observe that for any $x \in U$, we have $\mu(\langle \{x\} \rangle) \leq q \cdot \bar{\mu}(x)$. This is because every set in the support of μ has at most q elements, and then by definition $\bar{\mu}(x) = \sum_{S \in \langle x \rangle} \frac{\mu(S)}{|S|} \geq \frac{\mu(\langle x \rangle)}{q}$.

In addition, for every $T' \subseteq T$ we have $\mu(\langle T \rangle) \leq \mu(\langle T' \rangle)$, since in the left side of the inequality we sum up over (possibly) fewer sets. Hence, for $T' = \{x\}$ we get $\mu(\langle T \rangle) \leq \mu(\langle \{x\} \rangle)$. Together with the bound on $\bar{\mu}(x)$, this implies that:

$$\mu(\langle T \rangle) \leq \mu(\langle x \rangle) \leq q\bar{\mu}(x) \leq q \cdot \frac{k^{j-|T|}}{n} .$$

□

The proof of Claim 5.6 relies on the following combinatorial claim and a standard averaging argument.

Claim 5.9. *For any set $S \subseteq U$ with $|S| \leq q$, there are at least $c - q$ indices $j \in [c]$ for which B_j is a good boundary for S.*

Proof. Let $S \subseteq U$ with $|S| \leq q$. To prove the claim, we show that there are at most q "bad boundaries" for S, which implies that at least $c - q$ boundaries must be good. The proof uses a "token-shifting" argument. We conceptualize the elements of S as tokens distributed among the buckets. Initially, each bucket is assigned a number of tokens that is equal to the number of elements of S that are in this bucket. The total number of tokens is therefore $|S| \leq q$. We then apply an iterative redistribution process that shifts these tokens between buckets to produce a final configuration.

The core of the argument is to show that in this final configuration, if a bucket contains no tokens, then it must be a good boundary. We now proceed with the formal description of the redistribution process and its analysis.

The Token-Shifting Process. The process is defined as follows. We start with the initial token counts $w_j = |B_j \cap S|$ for $j \in [c]$. As long as there is any bucket B_j that contains two or more tokens ($w_j \geq 2$), we apply the following rule: move one token from bucket B_j to the bucket above it, B_{j+1}. That is, update the counts: $w_j \leftarrow w_j - 1$ and $w_{j+1} \leftarrow w_{j+1} + 1$. If $j = c$ and there is not bucket above it, only remove a token from B_c. The process ends when every bucket has one or zero token.

Loop Invariants. We claim that the following three invariants hold throughout the operation of the reassignment process.

1. The total number of tokens is at most $|S| \leq q$. That is, $\sum_{j=1}^{c} w_j \leq q$. This is because the total number of tokens starts at $|S|$ and can only decrease if bucket B_c is chosen.

2. If at some point a bucket has non-zero tokens, then it will never have zero tokens later. This is because we remove tokens from a bucket only if it has at least 2 tokens.

3. If bucket B_j has no tokens, then for any i, the total number of tokens in the i buckets below B_j is at least the number of elements of S in those buckets. That is, $\sum_{j'=j-i}^{j} w_{j'} \geq \sum_{j'=j-i}^{j} |B_{j'} \cap S|$. This is because a bucket with a non-zero weight will never become zero, and the total number of tokens in a sequence of buckets can decrease only when the bucket just above them gains a token.

Post-loop claims. From these invariants, we deduce the following claims about the weights at the end of the loop:

1. There are at most q non-zero buckets. Each bucket ends with a weight of at most 1, and by the first invariant the total number of tokens is at most q.

2. If a bucket B_j has no tokens, then it is a good boundary. Otherwise, there is some ℓ such that $\left|S \cap B_{[j-i,j]}\right| > i$. But according to the third invariant, since B_j contains no tokens, the total number of tokens in the i buckets below B_j is at least $\left|S \cap B_{[j-i,j]}\right| > i$. This is a contradiction, since after the loop ends each of these buckets has at most one token, and B_j contains no tokens.

Together, these claims imply that there are at least $c - q$ good boundaries for S. □

We can now complete the proof of Claim 5.6

Proof of Claim 5.6. Let $G(S) \subseteq [c]$ be the set of good boundaries for a set S. By Claim 5.9, $|G(S)| \geq c - q$ for every set S, and hence the expected number of good boundaries for a set $S \sim \mu$ is also lower bounded:

$$\mathbb{E}_{S\sim\mu}[|G(S)|] \geq c - q$$

On the other hand, by linearity of expectation, the expectation can be expressed as:

$$\begin{aligned}\mathbb{E}_{S\sim\mu}[|G(S)|] &= \mathbb{E}_{S\sim\mu}\left[\sum_{j=1}^{c} \mathbb{I}[j \in G(S)]\right] \\ &= \sum_{j=1}^{c} \mathbb{E}_{S\sim\mu}[\mathbb{I}[j \in G(S)]] \\ &= \sum_{j=1}^{c} \mu\left(\{S \mid j \in G(S)\}\right)\end{aligned}$$

Combining these facts implies there must exist an index $j^* \in [c]$ such that

$$\mu(\{S \mid j^* \in G(S)\}) \geq \frac{c-q}{c}\,.$$

We define the family $\mathcal{D}$ as:

$$\mathcal{D} = \{S \in \text{supp}(\mu) \mid B_{j^*} \text{ is a good boundary for } S\}$$

For this choice, we have:

$$\mu(\mathcal{D}) \geq \frac{c-q}{c} = 1 - \frac{q}{c}$$

□

Finally, we prove the simplest claim, Claim 5.8.

Proof of Claim 5.8. By the definition of the buckets, every element $u \in B_{[j+1,c]}$ has a weighted degree $\bar{\mu}(u) > \frac{k^j}{n}$. Summing this lower bound over all elements in $B_{[j+1,c]}$ yields:

$$\sum_{u\in B_{[j+1,c]}} \bar{\mu}(u) > \left|B_{[j+1,c]}\right| \cdot \frac{k^j}{n}\,.$$

On the other hand, since $\bar{\mu}$ is a probability distribution over the universe U, the sum of probabilities over any subset of U is at most 1:

$$\sum_{u\in B_{[j+1,c]}} \bar{\mu}(u) \leq \sum_{u\in U} \bar{\mu}(u) = 1\,.$$

Combining these two inequalities gives:

$$\left|B_{[j+1,c]}\right| \cdot \frac{k^j}{n} < 1\,,$$

and rearranging the terms finishes the proof. □

□

5.2 Large robust daisies

We next use the spread lemma for distributions (Lemma 4.6) to argue that if $\mu^{\circ K}$ is spread, then μ is a robust daisy with kernel K.

First, we make the following observation:

Observation 5.10. *Fix $k > 1$ and $m \in (0, 1]$. Let μ be a (m, k)-spread distribution over $\mathcal{P}(U)$ and let $\mathcal{D} \subseteq \mathrm{supp}(\mu)$. Then, $\mu_{\mathcal{D}}$ is $\left(\frac{m}{\mu(\mathcal{D})}, k\right)$-spread.*

Proof. Let $T \subseteq U$ be a non-empty set. By the definition of conditioned distribution, and by the spreadness of μ, we have:

$$\mu_{\mathcal{D}}(\langle T \rangle) \leq \frac{\mu(\langle T \rangle)}{\mu(\mathcal{D})} \leq \frac{1}{\mu(\mathcal{D})} \cdot \frac{m}{k^{|T|}}$$

□

Armed with this observation, we show that a distribution which is spread outside a kernel is also a robust daisy.

Lemma 5.11. *Fix $k > 1$ and $m \in (0, 1]$. Let μ be a distribution over $\mathcal{P}_{\leq q}(U)$. Suppose $\mu^{\circ K}$ is (m, k)-spread for some $K \subseteq U$. Then, for every $\alpha > 2q$, μ is a (p, ε)-robust daisy with kernel K, $p = \alpha/k$ and $\varepsilon = \exp\left(-\frac{\alpha}{8qm \log |\mathrm{supp}(\mu)|}\right)$.*

Proof. To prove that μ is a (p, ε)-robust daisy, we need to show that for every $\mathcal{D} \subseteq \mathrm{supp}(\mu)$, the family $\mathcal{D} \setminus K := \{S \setminus K \mid S \in \mathcal{D}\}$ is $(p, \varepsilon^{\mu(\mathcal{D})})$-satisfying.

By assumption, $\mu^{\circ K}$ is (m, k)-spread. Therefore, by Observation 5.10, the distribution $(\mu^{\circ K})_{\mathcal{D} \setminus K}$ is $\left(\frac{m}{\mu^{\circ K}(\mathcal{D} \setminus K)}, k\right)$-spread.[11]

Now, by Lemma 4.6, for every $\alpha > 2q$ we have that $\mathrm{supp}((\mu^{\circ K})_{\mathcal{D} \setminus K}) = \mathcal{D} \setminus K$ is (p, ε')-satisfying with $p = \alpha/k$ and

$$\varepsilon' = \exp\left(-\frac{\alpha}{8q(m/\mu^{\circ K}(\mathcal{D} \setminus K)) \log |\mathcal{D} \setminus K|}\right) \leq \exp\left(-\frac{\alpha \cdot \mu(\mathcal{D})}{8qm \log |\mathrm{supp}(\mu)|}\right) = \varepsilon^{\mu(\mathcal{D})} .$$

where the inequality follows from the observation that $\mu^{\circ K}(\mathcal{D} \setminus K) \geq \mu(\mathcal{D})$ (Observation 5.3) and since $|\mathrm{supp}(\mu)| \geq |\mathcal{D}| \geq |\mathcal{D} \setminus K|$. □

We can now prove the main result of this section.

Proof of Lemma 5.1. The lemma follows from the Punctured Spread Distribution Lemma (Lemma 5.4), combined with the fact that spreadness implies robustness (Lemma 5.11).

Let $c > q$. By Lemma 5.4, there exists $j \in [c]$, $\mathcal{D} \subseteq \mathcal{P}_{\leq q}(U)$ and a non-empty $K \subseteq U$ such that $(\mu_{\mathcal{D}})^{\circ K}$ is a $(m, n^{1/c})$-spread distribution, where:

$$m = \frac{q}{\mu(\mathcal{D})} \cdot \frac{n^{j/c}}{n}, \qquad |K| \leq \frac{n}{n^{j/c}}, \qquad \mu(\mathcal{D}) \geq 1 - \frac{q}{c} .$$

[11] We remark that, perhaps unintuitively, it is not always true that $(\mu^{\circ K})_{\mathcal{D} \setminus K} = (\mu_{\mathcal{D}})^{\circ K}$.

Suppose $\alpha > 2q$. Now, we apply Lemma 5.11 to conclude that $\mu_{\mathcal{D}}$ is a (p, ε)-robust daisy with $p = \alpha/k = \alpha n^{-1/c}$ and

$$\begin{aligned}
\varepsilon &= \exp\left(-\frac{\alpha}{8qm \log|\mathcal{D}|}\right) && (\mathrm{supp}(\mu_{\mathcal{D}}) = \mathcal{D}) \\
&= \exp\left(-\frac{\alpha \cdot \mu(\mathcal{D}) \cdot n}{8q \cdot q \cdot n^{j/c} \cdot \log|\mathcal{D}|}\right) && (\text{plugging in } m) \\
&\leq \exp\left(-\frac{\alpha\mu(\mathcal{D}) \cdot |K|}{8q^2 \log|\mathcal{D}|}\right) && (\text{since } |K| \leq \frac{n}{n^{j/c}}) \\
&\leq \exp\left(-\frac{\alpha(1-q/c)|K|}{8q^2 \log|\mathcal{D}|}\right). && (\text{since } \mu(\mathcal{D}) \geq 1 - q/c)
\end{aligned}$$

Finally, since $j \geq 1$, we have $|K| \leq \frac{n}{n^{j/c}} \leq n^{1-1/c}$, as required. □

6 RLDC lower bounds

This section proves the main result (Theorem 1) of the paper.

In Section 6.1, we provide a detailed overview that motivates and discusses the definition of robust daisies and the "global sampler" strategy. In Section 6.2, we formally prove that if a relaxed decoder possesses the structure of robust daisies, this implies a lower bound on its block length. Section 6.1 and Section 6.2 are independent of the previous sections, and we encourage readers unfamiliar with the work of [GL21] to read them first.

Finally, in Section 6.3, we prove the main result, by using the Robust Daisy Lemma to show that any relaxed decoder can be transformed into one that has the structure of robust daisies.

6.1 Overview

At its core, the proof is a compression-based, information-theoretic argument: one cannot recover k bits of information by observing fewer than k bits, except with small probability.

For the rest of the overview, let $C : \{0,1\}^k \to \{0,1\}^n$ be a q-query RLDC with a non-adaptive decoder, constant decoding radius δ, and soundness probability σ. We show that if the block length n is too small as a function of the message length k, it is possible to recover, with a high probability, the entire k-bit message by querying fewer than k bits of the corresponding codeword.

Global sampler. A *global sampler* is a probabilistic algorithm with oracle access to a *valid codeword* $C(x)$. It samples each bit of the codeword independently with some probability p, and its goal is to recover the entire message x. With a high probability, the sampler queries $\Theta(pn)$ bits in total. Therefore, if it succeeds in recovering the message x, it implies a bound of $pn = \Omega(k)$. For concreteness, a sampling probability of exactly $p = n^{-1/q}$ would give the bound $n^{1-\frac{1}{q}} = \Omega(k)$, which implies $n = \Omega\left(k^{1+\frac{1}{q-1}}\right)$.

The global sampler leverages the relaxed decoder of C to recover the bits of x. To recover the i-th bit, x_i, the global sampler aims to simulate the relaxed decoder for index i. The challenge is that the sampler's random samples must be sufficient to decode all k indices *simultaneously*, whereas the relaxed decoder for each index i may query an arbitrary set, according to a distribution depending on i.

Warm-up: satisfying set systems. Consider a non-adaptive relaxed decoder for a code $C : \{0,1\}^k \to \{0,1\}^n$. To decode x_i, the decoder performs as follows: it picks a set S according to some distribution μ_i, queries the indices in S, and computes its output *deterministically* based on its sampled *local view*.[12]

If the global sampler happens to draw a set of samples that contains a full set from the support of μ_i, it can recover x_i by invoking the decoder's logic on that local view. Since the global sampler's input is a valid codeword, the relaxed decoder is guaranteed to decode correctly.

This idea is captured by the well-known notion of a *Satisfying Set System*. In the following, $W \sim \mathrm{Bin}(U,p)$ denotes a random subset of U where each element is included independently with probability p.

Definition 6.1 (Satisfying set system, restating Definition 4.1)**.** *Let $\mathcal{F}$ be a family of sets over a universe U. We say that $\mathcal{F}$ is (p,ε)-satisfying if*

$$\Pr_{W \sim Bin(U,p)}[\exists S \in \mathcal{F}, S \subseteq W] \geq 1 - \varepsilon\,.$$

This suggests that our goal is to show that each family $\mathcal{F}_i := \mathrm{supp}(\mu_i)$ is (p,ε)-satisfying for a small ε (e.g., $\varepsilon \ll 1/k$, to allow for a union bound over all k indices). However, it is unreasonable to expect the query families $\mathcal{F}_i$ to satisfy such a strong property. The decoder's queries may be highly non-uniform; for instance, to decode index i, the decoder might always query index i itself — a common feature in known constructions. In this case, the global sampler must happen to sample index i to capture any full set $S \in \mathcal{F}_i$. Since this occurs with probability p, the sampler would effectively need to query almost all of $C(x)$ to recover all of x, yielding a trivial bound. We must therefore handle these "heavy indices" differently.

The actual global sampler: robust daisies. Our strategy is to *guess* the values of $C(x)$ at the few "heavy" indices instead of trying to sample them. Let us denote the set of these heavy indices by a kernel $K \subseteq [n]$. To obtain a full local view for a query set S, the global sampler no longer needs to sample all of S; it only needs to sample the "light" elements in $S \setminus K$.

We call the sets $\{S \setminus K \mid S \in \mathcal{F}_i\}$ the *petals*. The intuition is that it might be sufficient for the family of petals to be satisfying, rather than the family of full query sets.

If the guess for the heavy indices is correct, each local view we construct is consistent with the valid codeword $C(x)$. Therefore, since the decoder never errs on valid codewords, applying its logic to any of these views will yield the correct output bit.

But what if our guess for the bits in K is incorrect? In this case, the local views we construct are not consistent with $C(x)$, but rather with a corrupted version of it. The number of corruptions is at most $|K|$. If $|K|$ is smaller than the decoding radius of the code, then the decoder's soundness guarantee holds. This means that for any guess, at least a large fraction (i.e., σ, where "fraction" is measured with respect to the distribution μ_i) of the query sets will lead the decoder to output the correct symbol or the special rejection symbol $\perp$.

This leads to the actual algorithm for the global sampler. To recover each bit x_i:

1. Sample the codeword $C(x)$ by picking each index i.i.d. with probability p.

[12] We can assume that this computation is deterministic, since the decoder is non-adaptive and must never err on valid codewords.

2. Iterate over all $2^{|K|}$ possible guesses for the values of the bits at the heavy indices in K.
3. For each guess, identify all petals $S \setminus K$ that were fully contained in the initial sample.
4. For each such petal, form a complete local view using the sampled values and the current guess. Feed all these views to the decoder's logic.
5. If all these simulated local views result in the same output bit, output that bit for x_i. Otherwise, continue.

The algorithm is guaranteed to produce an output - for the correct guess on K. By the decoder's completeness, every local view consistent with $C(x)$ leads to the correct output bit, ensuring a consensus. An incorrect output for x_i can only occur if, for some incorrect guess, *all* sampled petals happen to correspond to query sets that mislead the decoder.

Conversely, for any guess, the algorithm avoids an error as long as at least one sampled petal corresponds to a "good" query set — one that would lead the decoder to output the correct bit or $\perp$. By the soundness guarantee, a large fraction of query sets are good. Therefore, we expect that our sampler will likely hit one of the many corresponding "good" petals.

This motivates our central definition of a *Robust Daisy*. These structures differ from naive satisfying set systems in two ways. First, they allow for a kernel K to be handled separately. Second, they require that *any* sufficiently large sub-family is also satisfying. This ensures that even if we restrict our attention to the "good" sets (which form a large subfamily), we are still guaranteed to sample one of their petals.

For simplicity, our definition requires this property for *all* sub-families, a stronger condition we can achieve without extra cost.

Definition 6.2. *(Robust daisy, restating Definition 1.1) A distribution μ over $\mathcal{P}(U)$ is a* (p, ε)-robust daisy *with kernel $K \subseteq U$, if, for every $\mathcal{D} \subseteq \mathrm{supp}(\mu)$, the family of petals $\mathcal{D} \setminus K := \{S \setminus K \mid S \in \mathcal{D}\}$ is $(p, \varepsilon^{\mu(\mathcal{D})})$-satisfying. That is, if:*

$$\Pr_{W \sim Bin(U,p)}[\exists\, S \in \mathcal{D}, S \subseteq K \cup W] \geq 1 - \varepsilon^{\mu(\mathcal{D})}\,.$$

From robust daisies to a global sampler. Now, suppose that for some $i \in [k]$, the query-set distribution of the relaxed decoder for index i, denoted μ_i, is a (p, ε_i)-robust daisy with kernel K_i. Recall that at least a σ fraction of the petals are "good" (i.e., they lead to a correct output or $\perp$). Hence, by the discussion above, for any specific guess of the kernel values, the probability that the global sampler *fails* to sample a "good" petal is at most ε_i^{σ}. By taking a union bound over all $2^{|K_i|}$ possible guesses, the probability that the global sampler fails to find a good petal for *any* of the guesses, and hence outputs a wrong bit for x_i, is at most $2^{|K_i|} \cdot \varepsilon_i^{\sigma}$.

Taking another union bound over all $i \in [k]$, and assuming that every μ_i is a robust daisy, gives that with probability at least $1 - \sum_{i \in [k]} \varepsilon_i^{\sigma} \cdot 2^{|K_i|}$ the global sampler recovers *all* of the k bits of x, without any mistakes. This brings us to the exact requirement we need from the robust daisy, and the formal proof of the reduction, which we present in Section 6.2.

However, what if the distributions μ_i are not robust daisies? A simple argument shows that it is enough to extract a robust daisy of large density from each μ_i. Namely, if we find $\mathcal{D} \subseteq \mathrm{supp}(\mu)$ such that $\mu_{\mathcal{D}}$ is a robust daisy, we can modify the relaxed decoder such that instead of sampling a set $S \sim \mu$ and querying the indices of S, the modified decoder samples a

set $S \sim \mu_{\mathcal{D}}$. Assuming that $\mu(\mathcal{D})$ is large enough compared to σ (specifically, if $\mu(\mathcal{D}) > 1-\sigma$), then the modified decoder has non-trivial soundness, and can be used by the global sampler. We give a formal proof of this argument in Section 6.3, and use it to complete the proof of Theorem 1.

6.2 Lower bound for RLDCs with structured decoders

In this section, we prove a lower bound for RLDCs with relaxed decoders whose query set distributions are robust daisies.

Lemma 6.3. *Let $C : \{0,1\}^k \to \Sigma^n$ be a non-adaptive (q, δ, σ)-RLDC, and let p satisfy $\frac{3\ln(n)}{n} < p < 1$. For each $i \in [k]$, let μ_i be the decoder's query distribution for index i.*

Assume that for every $i \in [k]$, μ_i is a (p, ε_i)-robust daisy with a kernel $K_i \subseteq [n]$ such that

$$|K_i| \leq \delta n \quad \text{and} \quad \varepsilon_i^{\sigma} \leq \frac{1}{3k \cdot |\Sigma|^{|K_i|}} \,. \tag{5}$$

Then $k \leq 2pn \cdot \log|\Sigma|$.

Proof. We give a formal description of the global sampler, $\mathcal{G}$, in Algorithm 1.

Notation. In the description of $\mathcal{G}$ we use the following notation. Let $i \in [k]$, and let $\mathcal{B}_i$ be the corresponding relaxed decoder with query set distribution μ_i. Let $\mathcal{D}_i = \mathrm{supp}(\mu_i)$. On a codeword y, the relaxed decoder samples a query set $S \sim \mu_i$, obtains y_S and outputs some deterministic function of S and y_S. We can assume that this function is deterministic since the decoder is non-adaptive and never errs on valid codewords. Let $f_i : \mathcal{D}_i \times \Sigma^q \to \{0,1\}$ be this deterministic function.

The global sampler. The global sampler $\mathcal{G}$ operates in two phases:

1. Query Phase: First $\mathcal{G}$ samples each coordinate of y independently with probability p. Formally, it samples the coordinates $W \sim \mathrm{Bin}([n], p)$. If $|W| \geq 2pn$, $\mathcal{G}$ outputs a random $\hat{x} \in \{0,1\}^k$.

   Otherwise, $\mathcal{G}$ queries all the bits in W from y to obtain $w = y_W$.

2. Decoding Phase: For each $i \in [k]$, $\mathcal{G}$ decodes x_i as follows. Let

   $$\mathcal{D}_i^{\text{sampled}} = \{S \in \mathcal{D}_i \mid S \setminus K_i \subseteq W\}$$

   be the query sets of $\mathcal{D}_i$ whose petals are fully sampled by $\mathcal{G}$. For any $\kappa \in \Sigma^{|K_i|}$ and $S \in \mathcal{D}_i^{\text{sampled}}$, let $a_{S,\kappa}$ be the assignment of the variables of S that is consistent with κ on K_i and with w on $S \setminus K_i$.

   $\mathcal{G}$ iterates over each assignment $\kappa \in \Sigma^{|K_i|}$ and does the following check on the relaxed decoder $\mathcal{B}$ with decoding function f_i:

   For all $S \in \mathcal{D}_i^{\text{sampled}}$ with corresponding assignment $a_{S,\kappa}$, does $\mathcal{B}$ output the same bit $b \in \{0,1\}$? If yes, then $\mathcal{G}$ sets $\hat{x}_i = b$.

   If $\mathcal{G}$ never sets x_i in these iterations, i.e., no such κ exists, then it sets $\hat{x}_i = 0$.

Algorithm 1 Global Sampler $\mathcal{G}$

Input: $\mathcal{D}_i, f_i, K_i \; \forall \; i \in [k]$, query access to $y \in \Sigma^n$.
Output: $\hat{x} \in \{0,1\}^k$

Sample $W \sim \mathrm{Bin}([n], p)$.
if $|W| \geq 2pn$ **then**
    Output a random $\hat{x} \sim \{0,1\}^k$.
end if
$w \leftarrow y_W$.
for $i \in [k]$ **do**
    $\mathcal{D} \leftarrow \mathcal{D}_i, K \leftarrow K_i, f \leftarrow f_i, \hat{x}_i \leftarrow 0$.
    $\mathcal{D}^{\text{sampled}} \leftarrow \{S \in \mathcal{D} \mid S \setminus K \subseteq W\}$.
    for $\kappa \in \Sigma^{|K|}$ **do**
        for $S \in \mathcal{D}^{\text{sampled}}$ **do**
            $a_{S,\kappa} \leftarrow (\kappa_{S \cap K}, w_{S \setminus K})$
        end for
        if $\exists \; b \in \{0,1\}, \; \forall \; S \in \mathcal{D}^{\text{sampled}}, f(S, a_{S,\kappa}) = b$ **then**
            $\hat{x}_i \leftarrow b$.
        end if
    end for
end for
return $\hat{x}$.

Analysis. We show that $\mathcal{G}$ succeeds, i.e., it outputs x correctly, with probability at least $1/2$. $\mathcal{G}$ fails if it either samples $\geq 2pn$ coordinates, or decodes x_i incorrectly for some $i \in [k]$.

The first kind of failure happens if $|W| \geq 2pn$ when $W \sim \mathrm{Bin}([n], p)$. By Chernoff's bound, $\Pr[|W| \geq 2pn] \leq \exp(-pn/3) \leq 1/n$, since $p \geq 3\ln(n)/n$ by assumption.

In Claim 6.4, we argue that for each $i \in [k]$, $\mathcal{G}$ fails to decode x_i with probability at most $|\Sigma|^{|K_i|} \cdot \varepsilon_i^{\sigma}$. Since, by the hypothesis, $\varepsilon_i < \exp\left(-\frac{\log(3k)+|K_i|\log(|\Sigma|)}{\sigma}\right)$, this failure probability is at most $1/3k$.

First, we complete our argument assuming this claim:

$$\begin{aligned}
\Pr_W[\mathcal{G} \text{ fails}] &= \Pr_W[\mathcal{G} \text{ fails and } |W| \geq 2pn] + \Pr_W[\mathcal{G} \text{ fails and } |W| \leq 2pn] \\
&\leq \Pr_W[|W| \geq 2pn] + \Pr_W[\; \exists \; i \in [k], \mathcal{G} \text{ fails to decode } x_i] \\
&\leq 1/n + \sum_{i \in [k]} \Pr_W[\mathcal{G} \text{ fails to decode } x_i] \\
&\leq 1/n + k \cdot 1/3k \leq 1/2 \; ,
\end{aligned}$$

Hence, $\mathcal{G}$ decodes all of x with probability at least $1/2$. Since it makes at most $2pn$ queries, this implies $k < 2pn \log |\Sigma|$.

We are left to prove Claim 6.4, which ensures that $\mathcal{G}$ decodes each x_i with high probability, which finishes the proof.

Claim 6.4. *For any $i \in [k]$, the global sampler decodes x_i with probability at least $1 - |\Sigma|^{|K_i|}\varepsilon_i^{\sigma}$.*

Proof. Fix $i \in [k]$. Let $\mathcal{A} := \mathcal{B}_i$ be the relaxed decoder with query set distribution $\mu := \mu_i$ which is a (p, ε)-robust daisy with kernel $K := K_i$ for $\varepsilon := \varepsilon_i$. Let $\mathcal{D} := \mathcal{D}_i$ and $f := f_i$.

Recall that $\mathcal{G}$ iterates over each $\kappa \in \Sigma^{|K|}$. In the case where $\mathcal{G}$ guesses κ correctly, i.e., $\kappa = y_K$, we argue that $\mathcal{G}$ sets $\hat{x}_i$ correctly with probability at least $1 - \varepsilon$. For every other guess, we argue that with probability at least $1-\varepsilon^\sigma$, the global sampler either sets $\hat{x}_i$ correctly or does not change $\hat{x}_i$. By the union bound we then conclude that $\mathcal{G}$ sets $\hat{x}_i$ correctly with probability at least $1 - |\Sigma|^{|K|}\varepsilon^\sigma$.

Correct guess. Suppose $\kappa = y_K$. We next show that $\mathcal{G}$ sets x_i correctly with probability at least $1 - \varepsilon$. For each $S \in \mathcal{D}^{\text{sampled}}$, $a_{S,\kappa}$ agrees with the codeword y. Since the relaxed decoder never errs on valid codewords, $f(S, a_{S,\kappa}) = x_i$ for all $S \in \mathcal{D}^{\text{sampled}}$. Therefore, as long as $|\mathcal{D}^{\text{sampled}}| \geq 1$, $\mathcal{G}$ sets x_i correctly. Since μ is a (p, ε)-robust daisy and $\mathcal{D} = \text{supp}(\mu)$, this occurs with probability at least $1 - \varepsilon$.

Incorrect guess. Conversely, suppose that $\kappa \neq y_K$. We show that with probability at least $1 - \varepsilon^\sigma$, $\mathcal{G}$ does not set x_i incorrectly.

Let $y' \in \{0,1\}^n$ be the string that agrees with the guess κ on K, and agrees with y on $[n] \setminus K$. Now, each $a_{S,\kappa}$ is consistent with y', and $\text{dist}(y, y') \leq \frac{|K|}{n} \leq \delta$. Therefore, due to the soundness of the decoder:

$$\Pr_{S\sim\mu}[f(S, a_{S,\kappa}) \in \{x_i, \perp\}] \geq \sigma \ .$$

For any $\kappa \in \{0,1\}^K$, let $\mathcal{D}^{\text{good}}(\kappa) \subseteq \mathcal{D}$ be the query sets on which $\mathcal{A}$ decodes correctly when K is assigned κ, i.e., $\mathcal{D}^{\text{good}}(\kappa) = \{S \in \mathcal{D} \mid f(S, a_{S,\kappa}) \in \{x_i, \perp\}\}$, and note that the equation above implies $\mu(\mathcal{D}^{\text{good}}(\kappa)) \geq \sigma$.

Since μ is a (p, ε)-robust daisy and $\mathcal{D}^{\text{good}}(\kappa) \subseteq \mathcal{D}$, $\mathcal{D}^{\text{sampled}}$ contains a set from $\mathcal{D}^{\text{good}}(\kappa)$ with probability at least $1 - \varepsilon^\sigma$. Hence, in this case, there exists $S \in \mathcal{D}^{\text{good}}(\kappa) \cap \mathcal{D}^{\text{sampled}}$, which ensures that $\mathcal{G}$ does not set x_i incorrectly.

□

□

6.3 Lower bound for arbitrary RLDCs

In this section, we use the robust daisy lemma (Lemma 5.1) to show that an arbitrary RLDC can be transformed into an RLDC with a decoder whose query set distribution is a robust daisy with slightly worse soundness error.

Lemma 6.5. *Let $C : \{0,1\}^k \to \Sigma^n$ be an error correcting code with a non-adaptive (q, δ, σ)-relaxed decoder.*

Then, for every integer $c > \frac{q}{\sigma}$ such that $n^{-1/c} \leq \delta$, the code C also has a non-adaptive (q, δ, σ')-relaxed decoder where $\sigma' = \frac{c\sigma - q}{c - q}$, and for every $i \in [k]$, the query distribution μ_i of the new decoder for index i is a (p, ε_i)-robust daisy with kernel $K_i \subseteq [n]$ with

$$p = \frac{8q^3 \cdot \log 3k \cdot \log n \cdot \log |\Sigma|}{\sigma - q/c} \cdot n^{-1/c}$$

such that

$$|K_i| \leq \delta n, \qquad\qquad \varepsilon_i^{\sigma'} \leq \frac{1}{3k \cdot |\Sigma|^{|K_i|}} \ .$$

Proof. We begin by describing our modified relaxed decoder.

The modified decoder. For every $i \in [k]$, let $\mathcal{B}_i$ be the original (q, δ, σ)-relaxed decoder for C for index i, with query set distribution μ_i.

Let $c > q/\sigma$ be an integer such that $n^{-1/c} \leq \delta$, and let

$$\alpha := \frac{8q^3 \cdot \log 3k \cdot \log n \cdot \log |\Sigma|}{\sigma - q/c} .$$

By Lemma 5.1, there exists a subset $\mathcal{D}_i \subseteq \mathrm{supp}(\mu_i)$ such that $\mu_{\mathcal{D}_i}$ is a (p, ε_i)-robust daisy with non-empty kernel K_i, where:

$$p = \alpha n^{-1/c} \qquad \varepsilon_i = \exp\left(-\frac{\alpha(1-q/c)}{8q^2 \log |\mathcal{D}_i|} \cdot |K_i|\right) \qquad |K_i| \leq n^{1-1/c} \qquad \mu(\mathcal{D}_i) \geq 1 - q/c \, . \tag{6}$$

The modified decoder $\mathcal{A}$ works as follows. For each $i \in [k]$, the decoder for index i samples a set $S \sim \mu_{\mathcal{D}_i}$. It then decodes x_i according to the deterministic function f_i of $\mathcal{B}_i$ restricted to $\mathcal{D}_i$.

Since $\mathcal{D}_i \subseteq \mathrm{supp}(\mu_i)$, every set it samples is also a set the original decoder could have sampled. Therefore, the new decoder also makes at most q queries, and never errs on valid codewords.

Soundness. Recall that $\sigma' := \frac{c\sigma - q}{c - q}$. We next show that the soundness probability of $\mathcal{A}$ is at least σ'.

That is, we need to show that for any $x \in \{0,1\}^k$ and $w \in \Sigma^n$ such that $\mathrm{dist}(w, C(x)) \leq \delta$, we have

$$\Pr_{S \sim \mu_{\mathcal{D}_i}}[\mathcal{A}^w(i) = x_i] \geq \frac{c\sigma - q}{c - q}. \tag{7}$$

Now,

$$\begin{aligned} 1 - \sigma &\geq \Pr_{S \sim \mu_i}[\mathcal{B}^w(i) \neq x_i] \\ &= \Pr_{S \sim \mu_{\mathcal{D}_i}}[\mathcal{B}^w(i) \neq x_i]\mu(\mathcal{D}_i) + \Pr_{S \sim \mu_{\overline{\mathcal{D}_i}}}[\mathcal{B}^w(i) \neq x_i]\mu(\overline{\mathcal{D}_i}) \\ &\geq \Pr_{S \sim \mu_{\mathcal{D}_i}}[\mathcal{B}^w(i) \neq x_i](1 - q/c) = \Pr_{S \sim \mu_{\mathcal{D}_i}}[\mathcal{A}^w(i) \neq x_i](1 - q/c) \end{aligned}$$

where $\overline{\mathcal{D}_i} = \mathrm{supp}(\mu_i) \setminus \mathcal{D}_i$ and using $\mu(\mathcal{D}_i) \geq 1 - q/c$. Rearranging the inequality gives Equation (7).

We conclude that $\mathcal{A}$ is a (q, δ, σ')-relaxed decoder for C.

Parameter verification. By Equation (6), for each $i \in [k]$ the query set distribution $\mu_{\mathcal{D}_i}$ is a (p, ε_i)-robust daisy with kernel K_i where:

$$p = \alpha n^{-1/c} = \frac{8q^3 \log n \log 3k \log |\Sigma|}{\sigma - q/c} \cdot n^{-1/c} .$$

By assumptions, $|K_i| \leq n^{1-1/c} \leq \delta n$.

Since each set in $\mathcal{D}_i$ has at most q elements, we have $|\mathcal{D}_i| \leq n^q \implies q \log n \geq \log |\mathcal{D}_i|$. Plugging the definition of σ' in the choice of α, we infer the following bound:

$$\alpha = \frac{8q^3 \log n \log 3k \log |\Sigma|}{\sigma - q/c} \geq \frac{8q^2 \log |\mathcal{D}_i|}{\sigma'(1 - q/c)} \cdot \log(3k) \cdot \log |\Sigma|,$$

and then:

$$\varepsilon_i^{\sigma'} = \exp\left(-\frac{\alpha \cdot (1 - q/c)}{8q^2 \log |\mathcal{D}_i|} \cdot |K_i| \cdot \sigma'\right) \leq \exp\left(-\log(3k) \cdot |K_i| \cdot \log |\Sigma|\right) \leq \frac{1}{3k \cdot |\Sigma|^{|K_i|}},$$

as required. □

Finally, we combine Lemma 6.5 and Lemma 6.3 to derive our main theorem.

Proof of Theorem 1. We prove that:

$$\frac{k}{\log^2 k} \leq 38q^4\sigma^{-2} \cdot \log^2 |\Sigma| \cdot n^{1 - \frac{1}{\lceil \frac{q}{\sigma} \rceil + 1}}. \tag{8}$$

This yields the statement of Theorem 1 after rearrangement of the terms.

Let $c := \lceil \frac{q}{\sigma} \rceil + 1$. Note that $q/\sigma < c < 2q/\sigma$, and therefore, $n^{-1/c} \leq n^{-\frac{\sigma}{2q}} \leq \delta$.

By Lemma 6.5, C has a non-adaptive (q, δ, σ')-relaxed decoder $\mathcal{B}$ with $\sigma' = \frac{c\sigma - q}{c - q}$ where for every $i \in [k]$, the query distribution μ_i of $\mathcal{B}$ on input i is a (p, ε_i)-robust daisy with a non-empty kernel $K_i \subseteq [n]$ that satisfies Equation (5) with p set according to the statement of Lemma 6.5.

Note that $\sigma - q/c = \sigma - \frac{q}{\lceil q/\sigma \rceil + 1} \geq \sigma - \frac{q}{q/\sigma + 1} = \frac{\sigma^2}{\sigma + q} \geq \frac{\sigma^2}{q+1}$. Furthermore, we can assume without loss of generality that $k \geq n^{1 - \frac{1}{q+1}}$, since otherwise we are already done. Therefore, $\log n \leq \frac{(q+1)\log k}{q} \leq 2 \log k$ (since $q \geq 2$). Combining these two inequalities,

$$\frac{8q^3 \cdot \log 3k \cdot \log n \cdot \log |\Sigma|}{\sigma - q/c} \leq 8q^3 \cdot (\log 3 + \log k) \cdot \frac{q+1}{\sigma^2} \cdot 2 \log k \cdot \log |\Sigma| < 19q^4\sigma^{-2} \log^2 k \log |\Sigma|.$$

Now, since $\mathcal{B}$ satisfies the guarantees of Equation (5) with

$$p = \frac{8q^3 \cdot \log 3k \cdot \log n \cdot \log |\Sigma|}{\sigma - q/c} \cdot n^{-1/c} < 19q^4\sigma^{-2} \log^2 k \cdot \log |\Sigma| \cdot n^{-1/c}$$

we can apply Lemma 6.3 (and note that $p > \frac{3 \ln n}{n}$ by the choice of parameters) to conclude:

$$k < 2pn \log |\Sigma| < 38q^4\sigma^{-2} \log^2 k \cdot \log^2 |\Sigma| \cdot n^{1 - \frac{1}{\lceil \frac{q}{\sigma} \rceil + 1}}.$$

□

By rearrangement of the terms in Equation (8), we derive a query lower bound for RLDCs with constant rate.

Corollary 6.6 (Constant rate)**.** *Fix a constant $\sigma > 1/2$ and let $\delta > n^{-1/4q}$. Let $C : \{0,1\}^k \to \{0,1\}^n$ be a non-adaptive (q, δ, σ)-RLDC, and suppose that $n = O(k)$. Then, $q = \Omega\left(\frac{\log k}{\log \log k}\right)$.*

Proof. Suppose $n < d \cdot k$ for some constant $d > 0$. Note that $\lceil q/\sigma \rceil + 1 < 2q + 2$. By Theorem 1, we know that

$$\frac{k}{\log^2 k} \leq 38q^4\sigma^{-2} \cdot n^{1-\frac{1}{\lceil \frac{q}{\sigma} \rceil + 1}} < q^5 \cdot d \cdot k^{1-\frac{1}{2q+2}} < q^6 \cdot k^{1-\frac{1}{2q+2}}$$

where we use that $q = \omega(1)$ (otherwise it is easy to verify that $n = \omega(k)$) and $\sigma > 1/2$.

This implies that

$$k^{\frac{1}{2q+3}} \leq \frac{k^{\frac{1}{2q+2}}}{\log^2 k} \leq q^6$$

where the first inequality follows for a large enough k. We conclude that

$$\log k < 6(2q+3)\log q,$$

which directly implies the stated bound. □

Our lower bound for non-adaptive RLDCs extends to the important case of linear RLDCs, by using a known reduction.

Remark 6.7. *[Gol24a] showed that any linear (q, δ, σ)-RLDC can be turned into $(q+1, \delta, \sigma)$-query non-adaptive RLDCs. We use this transformation along with Theorem 1 and Corollary 6.6 to obtain Corollary 2 and Corollary 3.*

Acknowledgments

Tom Gur thanks Victor Seixas Souza and Marcel Dall'Agnol for insightful conversations about t-daisies. Sidhant Saraogi thanks Alexander Golovnev and Chao Yan for helpful discussions. Guy Goldberg thanks Irit Dinur, Oded Goldreich, and Guy Rothblum for valuable discussions about relaxed LDCs.

References

[AGKM22] Omar Alrabiah, Venkatesan Guruswami, Pravesh Kothari, and Peter Manohar. A near-cubic lower bound for 3-query locally decodable codes from semirandom CSP refutation. *Electron. Colloquium Comput. Complex.*, TR22-101, 2022. 3

[ALWZ21] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795 – 815, 2021. 4, 9

[AS16] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley Publishing, 4th edition, 2016. 13

[AS21] Vahid R. Asadi and Igor Shinkar. Relaxed Locally Correctable Codes with Improved Parameters. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:12, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 1, 3

[BCW21] Tolson Bell, Suchakree Chueluecha, and Lutz Warnke. Note on sunflowers. *Discrete Mathematics*, 344(7):112367, 2021. 8

[BGH+06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil Vadhan. Robust pcps of proximity, shorter pcps, and applications to coding. *SIAM Journal on Computing*, 36(4):889–974, 2006. 1, 3, 13

[BHKL25] Arpon Basu, Jun-Ting Hsieh, Pravesh K. Kothari, and Andrew D. Lin. Improved lower bounds for all odd-query locally decodable codes. In *FOCS*, 2025. 3

[CGR+25] Bruno Cavalar, Mika Göös, Artur Riazanov, Anastasia Sofronova, and Dmitry Sokolov. Monotone circuit complexity of matching. *arXiv:2507.16105*, 2025. 4

[CGS22] Alessandro Chiesa, Tom Gur, and Igor Shinkar. Relaxed locally correctable codes with nearly-linear block length and constant query complexity. *SIAM Journal on Computing*, 51(6):1839–1865, 2022. 1, 3

[CKR22] Bruno Pasqualotto Cavalar, Mrinal Kumar, and Benjamin Rossman. Monotone circuit lower bounds from robust sunflowers. *Algorithmica*, 2022. 4

[CY22] Gil Cohen and Tal Yankovitz. Relaxed locally decodable and correctable codes: Beyond tensoring. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 24–35, 2022. 1

[CY24] Gil Cohen and Tal Yankovitz. Asymptotically-Good RLCCs with $\log^{2+o(1)} n$ Queries. In Rahul Santhanam, editor, *39th Computational Complexity Conference (CCC 2024)*, volume 300 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 8:1–8:16, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 1, 3, 4

[DGL23] Marcel Dall'Agnol, Tom Gur, and Oded Lachish. A structural theorem for local algorithms with applications to coding, testing, and verification. *SIAM Journal on Computing*, 52(6):1413–1463, 2023. 1, 2, 4, 5, 6

[DGMT22] Marcel Dall'Agnol, Tom Gur, Subhayan Roy Moulik, and Justin Thaler. Quantum proofs of proximity. *Quantum*, 6:834, 2022. 1

[Efr12] Klim Efremenko. 3-query locally decodable codes of subexponential length. *SIAM Journal on Computing*, 41(6):1694–1703, 2012. 1

[FKNP21] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. Thresholds versus fractional expectation-thresholds. *Ann. Math.*, 194(2):475, September 2021. 4

[GG18] Oded Goldreich and Tom Gur. Universal locally testable codes. *Chicago Journal OF Theoretical Computer Science*, 3:1–21, 2018. 1

[GG21] Oded Goldreich and Tom Gur. Universal locally verifiable codes and 3-round interactive proofs of proximity for csp. *Theoretical computer science*, 878:83–101, 2021. 1

[GGK19] Oded Goldreich, Tom Gur, and Ilan Komargodski. Strong locally testable codes with relaxed local decoders. *ACM Trans. Comput. Theory*, 11(3), April 2019. 1

[GL21] Tom Gur and Oded Lachish. On the power of relaxed local decoding algorithms. *SIAM Journal on Computing*, 50(2):788–813, 2021. 1, 2, 3, 4, 5, 6, 7, 9, 24

[Gol23] Oded Goldreich. On the lower bound on the length of relaxed locally decodable codes. *Electron. Colloquium Comput. Complex.*, TR23-064, 2023. 2, 4, 5, 6, 9

[Gol24a] Guy Goldberg. Linear Relaxed Locally Decodable and Correctable Codes Do Not Need Adaptivity and Two-Sided Error. In Karl Bringmann, Martin Grohe, Gabriele Puppis, and Ola Svensson, editors, *51st International Colloquium on Automata, Languages, and Programming (ICALP 2024)*, volume 297 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 74:1–74:20, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 1, 2, 4, 5, 32

[Gol24b] Oded Goldreich. On the relaxed LDC of BGHSV: A survey that corrects the record. *Electron. Colloquium Comput. Complex.*, TR24-078, 2024. 1, 3

[GR18] Tom Gur and Ron D Rothblum. Non-interactive proofs of proximity. *Computational Complexity*, 27(1):99–207, 2018. 1

[GRR20] Tom Gur, Govind Ramnarayan, and Ron Rothblum. Relaxed locally correctable codes. *Theory of Computing*, 16(18):1–68, 2020. 1, 3

[JM25] Oliver Janzer and Peter Manohar. A $k^{q/(q-2)}$ lower bound for odd query locally decodable codes from bipartite Kikuchi graphs. In *FOCS*, 2025. 3

[KdW03] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. In *STOC*, pages 106–115. ACM, 2003. 3

[KM24] Vinayak M. Kumar and Geoffrey Mon. Relaxed local correctability from local testing. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1585–1593, New York, NY, USA, 2024. Association for Computing Machinery. 1, 3

[KS17] Swastik Kopparty and Shubhangi Saraf. Local testing and decoding of high-rate error-correcting codes. *Electron. Colloquium Comput. Complex.*, TR17-126, 2017. 3

[KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, STOC '00, page 80–86, New York, NY, USA, 2000. Association for Computing Machinery. 3

[LSZ20] Shachar Lovett, Noam Solomon, and Jiapeng Zhang. From dnf compression to sunflower theorems via regularity. In *Proceedings of the 34th Computational Complexity Conference*, CCC '19, 2020. 4

[MNWSZ25] Elchanan Mossel, Jonathan Niles-Weed, Nike Sun, and Ilias Zadik. A bayesian proof of the spread lemma. *Random Struct. Algorithms*, 66(4), July 2025. 7

[PP23] Jinyoung Park and Huy Pham. A proof of the Kahn–Kalai conjecture. *J. Amer. Math. Soc.*, 37(1):235–243, aug 2023. 4

[Rao20] Anup Rao. Coding for Sunflowers. *Discrete Analysis*, feb 25 2020. 4, 8

[Rao25] Anup Rao. The story of sunflowers. *arXiv preprint arXiv:2509.14790*, 2025. 4

[Ros14] Benjamin Rossman. The monotone complexity of k-clique on random graphs. *SIAM Journal on Computing*, 43(1):256–279, 2014. 2, 4, 9

[Tal10] Michel Talagrand. Are many small sets explicitly small? In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 13–36, New York, NY, USA, 2010. Association for Computing Machinery. 4, 7

[Woo07] David P. Woodruff. New lower bounds for general locally decodable codes. *Electron. Colloquium Comput. Complex.*, TR07-006, 2007. 3

[Woo12] David P. Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. *J. Comput. Sci. Technol.*, 27(4):678–686, 2012. 3

[Yek08] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *J. ACM*, 55(1), February 2008. 1

[Yek12] Sergey Yekhanin. Locally decodable codes. *Found. Trends Theor. Comput. Sci.*, 6(3):139–255, 2012. 3

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il