Electronic Colloquium on Computational Complexity, Report No. 293 (2026)

Lower Bounds for Adaptive Relaxed Locally Decodable Codes via the Capped Horvitz–Thompson Estimator

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

Tom Gur†
University of Cambridge
tg508@cam.ac.uk

Sidhant Saraogi‡
Georgetown University
ss4456@georgetown.edu

November 3, 2026

Abstract

We prove a nearly optimal lower bound on the blocklength of general relaxed locally decodable codes (RLDCs), extending the result of [GGS26] to nonlinear encodings and adaptive decoders with two-sided error. Namely, we show that any q-query RLDC $C\colon \{0,1\}^k \to \{0,1\}^n$ must satisfy $n = k^{1+\Omega(1/q)}$, closely matching the upper bound of $n = k^{1+O(1/q)}$ by Ben-Sasson, Goldreich, Harsha, Sudan, and Vadhan (STOC 2004).

Our proof reduces showing RLDC lower bounds to estimating the output distribution of adaptive decoders. To this end, we introduce a capped Horvitz–Thompson estimator with uniformly bounded discrete derivatives on the p-biased cube. We establish its concentration of measure via self-bounding inequalities and control the downward bias induced by capping through a spreadness argument.

*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

The notion of relaxed locally decodable codes (RLDCs) is a well-studied relaxation of locally decodable codes, which allows the decoder to abort on corrupted inputs while still requiring it to correctly decode valid codewords.

More precisely, a q-RLDC is an error-correcting code $C : \{0,1\}^k \to \{0,1\}^n$ that admits a randomized decoder D that recovers individual message bits using at most q queries to a purported codeword $w \in \{0,1\}^n$. On a valid codeword $w = C(x)$ and index i, the decoder returns x_i with probability at least $2/3$. On a noisy codeword within the decoding radius, it must either return x_i or an abort symbol $\perp$ with probability at least $2/3$.

In a seminal work, Ben-Sasson, Goldreich, Harsha, Sudan, and Vadhan [BGH+06] introduced this relaxation and obtained q-RLDCs with blocklength $n = k^{1+O(1/q)}$ (see also [Gol24b]). This implies $O(1)$-RLDCs with nearly linear length, whereas the best known $O(1)$-query locally decodable codes have superpolynomial length [Yek08, Efr12]. Despite receiving much attention [GR18, GG18, GGK19, GRR20, AS21, GG21, DGMT22, CGS22, CY22, Gol24b, KM24, CY24, GMWZ26], this original construction still achieves the best parameters for $O(1)$-RLDCs.

This lack of progress is partially explained by recent lower bounds. For constant-query RLDCs, Dall'Agnol, Gur, and Lachish [DGL23] (building on [GL21]) have shown the general lower bound $n = k^{1+\widetilde{\Omega}(1/q^2)}$. More recently, a nearly optimal lower bound of $n = k^{1+\Omega(1/q)}$ was shown for RLDCs with *non-adaptive decoders* and perfect completeness [GGS26]. Using the reduction of [Gol24a], this bound also extends to all linear RLDCs.

Nevertheless, the question of the optimal lower bound for general RLDCs remained open. In particular, it was unknown whether adaptivity and non-linearity allow q-query RLDCs of length $n = k^{1+o(1/q)}$.

In this work, we resolve this question by proving a nearly optimal lower bound on the blocklength of *any* RLDC.

Theorem 1. *Suppose $C : \{0,1\}^k \to \{0,1\}^n$ is a q-RLDC with constant decoding radius. Then, for every constant $\varepsilon > 0$,*

$$n = \Omega_\varepsilon\left(q^{-2}\left(\frac{k}{\log k}\right)^{1+\frac{1}{(2+\varepsilon)q}}\right).$$

This recovers the nearly optimal lower bound in [GGS26], without assuming either non-adaptivity or perfect completeness. In particular, in the constant-query regime, Theorem 1 gives a bound of $n = k^{1+\Omega(1/q)}$, matching the exponent of the upper bound in [BGH+06] up to constant factors. In the constant-rate regime, this implies a lower bound of $q = \Omega(\log k / \log\log k)$. The result extends to completeness and soundness probabilities other than $2/3$. See Theorem 5.1 for the precise statement.

1.1 The Capped Horvitz–Thompson Estimator

The proof of Theorem 1 introduces the notion of the *capped Horvitz–Thompson Estimator*, which may be of independent interest. As we discuss in Section 2, our approach to proving the adaptive RLDC lower bound relies on estimating the total mass of certain sub-distributions over query sets, using only binomially sampled input coordinates.

More precisely, consider a sub-distribution μ over subsets of $[n]$. Let $W \subseteq [n]$ be obtained by sampling each coordinate independently with probability p. We would like to estimate the total mass $\sum_S \mu(S)$, including the contribution of non-sampled sets $S \not\subseteq W$, using only the weights $\mu(S)$ of sets $S \subseteq W$. In our application, this total mass corresponds to the probability that the decoder produces a specified output on a specified input string.

The *Horvitz–Thompson estimator* [HT52] is a classical statistical tool used to obtain an unbiased estimate of the total mass by assigning each sampled set S the weight of $\mu(S)p^{-|S|}$, compensating for its inclusion probability $p^{|S|}$. Although unbiased, this estimator need not concentrate around its mean, as overlapping sets can produce highly dependent contributions, and a single sampled coordinate may have a large influence on the estimate.

To address this, we impose a cap on the marginal contribution of each coordinate. The resulting Γ-capped Horvitz–Thompson (CHT) estimator bounds the discrete derivatives so that adding or removing one coordinate from the sample changes the estimate by at most Γ, while preserving as much of the original weight as these constraints permit. In the following definition we identify the sampled set W with its indicator vector $x = \mathbf{1}_W \in \{0,1\}^n$.

Definition 1.1. *Let μ be a sub-distribution over subsets of $[n]$, and fix $p \in (0,1]$ and $\Gamma > 0$. For $x \in \{0,1\}^n$, write $x_S = \prod_{i \in S} x_i$. The Horvitz–Thompson estimator is given by*

$$R_{\mu,p}(x) = \sum_{S \subseteq [n]} \mu(S) p^{-|S|} x_S.$$

The Γ-Capped p-Horvitz–Thompson estimator $\overline{R}_{\mu,p,\Gamma}(x)$ *is the maximum total weight* $\sum_{S \subseteq [n]} w_S$ *over assignments of weights*

$$0 \le w_S \le \mu(S) p^{-|S|} x_S \ ,$$

satisfying $\sum_{S \ni i} w_S \le \Gamma$ *for every* $i \in [n]$.

Using concentration inequalities for self-bounding functions, we show that the capped Horvitz–Thompson estimator concentrates around its mean, with bounds depending on the capping parameter Γ. The tension is, however, that the capping also induces a downward bias of the estimator, which can be large for general sub-distributions. We show that if the underlying sub-distribution is *spread*, then this bias is small. Together, these results show that the CHT estimator concentrates around its mean for any sub-distribution, and that this mean is close to the total mass when the sub-distribution is sufficiently spread.

An analytic perspective. The Horvitz–Thompson estimator can be viewed as a multilinear polynomial with nonnegative coefficients. The CHT estimator can be interpreted as the pointwise maximum of such polynomials whose derivatives are uniformly bounded. The downward bias is therefore exactly the ℓ_1 distance between the two functions on the p-biased cube. We believe this perspective suggests interesting connections to Fourier analysis on the p-biased cube, which we leave for future work.

1.2 Related Work

While the state of the art of constant-query RLDCs is still the original work of Ben-Sasson, Goldreich, Harsha, Sudan, and Vadhan [BGH+06], which was observed in [Gol24b] to achieve blocklength $n = k^{1+O(1/q)}$, much work has been done on constructions of RLDCs and closely related notions. This includes RLDCs that are also locally testable [GR18, GGK19], and

relaxed locally correctable codes (RLCCs) [GRR20, CGS22, AS21, CY22], where the decoder recovers individual *codeword* symbols. In the constant-rate regime, Kumar and Mon [KM24] constructed RLCCs with polylogarithmic query complexity. Cohen and Yankovitz [CY24] subsequently optimized the query complexity to $(\log n)^{2+o(1)}$. These constructions are linear, and hence also yield RLDCs, leaving a gap to our lower bound of $q = \Omega(\log k / \log\log k)$.

In terms of lower bounds, our proof follows the sample-based decoding approach underlying previous RLDC lower bounds [GL21, DGL23, Gol23, GGS26]. Goldberg [Gol24a] showed that for the special case of linear codes, it is possible to convert arbitrary decoders to non-adaptive ones with perfect completeness, with only minor overhead in query complexity. Consequently, the nearly tight lower bound of [GGS26] applies both to non-adaptive RLDCs with perfect completeness and to general linear RLDCs.

AI Disclosure

The main conceptual and technical ideas in this paper are due to the authors. We used ChatGPT (GPT-6 Astra and GPT-5.6 Sol) in proofreading, literature review, minor stylistic editing, and assistance with certain technical claims. More specifically, we used it in the development of Lemma 4.15, where it helped us derive a first-principles bound on the expectation gap, instead of relying on the Kim–Vu inequality, which would incur an additional $\log^{2q} k$ loss in the resulting lower bound. We also used it in deriving the proof of the technical lemma Lemma 4.14, and in working out details of the counterexample in Section 2.3. The authors verified the correctness and originality of all content in this work.

2 Proof Overview

Our starting point is the *sample-based decoder* paradigm for showing RLDC lower bounds. We first review this paradigm and the obstacles that arise for adaptive decoders with two-sided error, and then describe how our approach overcomes them using the Capped Horvitz–Thompson estimator.

2.1 From Sample-Based Decoders to Lower Bounds

The core of the lower bound proof is an information-theoretic argument. By Fano's inequality, recovering k independent bits, each with error probability at most $\varepsilon < 1/2$ requires observing $(1 - h_2(\varepsilon))k = \Omega(k)$ bits of information.

In this work, we introduce a new notion of a *sample-based decoder*. Such a decoder gets sample access to a valid codeword $C(x)$, and it must recover each bit of x with error probability at most $\varepsilon < 1/2$. Since the sample-based decoder observes, in expectation, pn codeword symbols, the argument above implies that if a code C has a sample-based decoder with sampling probability p, then $pn = \Omega(k)$.

More precisely, let X be uniformly distributed over $\{0,1\}^k$, and let $W \subseteq [n]$ be the random set of sampled coordinates. Intuitively, recovering each bit with error probability at most ε means that after the sample-based decoder observes $(W, C(X)|_W)$, at most $kh_2(\varepsilon)$ bits of uncertainty remain about X. More precisely, we use Fano's inequality and the subadditivity of conditional entropy to prove that $H(X \mid W, C(X)|_W) \le kh_2(\varepsilon)$. On the other hand, since $H(X) = k$, observing an expected number of pn codeword bits leaves at least $k - pn$ bits of

uncertainty about X. Namely, $H(X \mid W, C(X)|_W) \geq k - np$. Combining the bounds yields the result.

This global-decoding paradigm has been used in all previous RLDC lower bounds. However, previous work imposed the stronger requirement that the decoder must recover all of the bits of x *simultaneously* with high probability. We view the introduction of this weaker notion of a sample-based decoder as a conceptual contribution, as it is the natural information-theoretic formulation of sample-based decoding.

Besides isolating the minimal guarantee needed for the lower-bound argument, it allows us to further improve the parameters of the lower bound of [GGS26] and simplify the logic of the sample-based decoder. Specifically, ensuring simultaneous recovery of all k message bits via a union bound requires reducing the failure probability for each bit to $O(1/k)$, incurring an additional logarithmic factor in the sampling probability. In contrast, our coordinate-wise guarantee avoids union bound and the associated logarithmic factor. This improves the base of the exponent in the lower bound from $\frac{k}{\log^2 k}$ in [GGS26] to $\frac{k}{\log k}$ in this work.

2.2 The Non-Adaptive, 1-Sided Error Case

We begin by recalling the sample-based decoding strategy of [GL21, GGS26].

Consider a relaxed decoder for a code $C : \{0,1\}^k \to \{0,1\}^n$. If it is non-adaptive, and never errs on valid codewords, then without loss of generality it can be fully described by a set of distributions $\{\mu_i\}_{i\in[k]}$, each over subsets of $[n]$. To decode x_i from a corrupted codeword $y \in \{0,1\}^n$, the relaxed decoder samples a set S according to μ_i and queries the indices in S. It then outputs according to the local view $y|_S$: if there exists $x \in \{0,1\}^k$ consistent with the local view (that is, $C(x)|_S = y|_S$), it outputs x_i, and otherwise it outputs $\perp$.[1]

If the sample-based decoder samples all the indices of some set S from the support of μ_i, it can recover x_i with certainty by invoking the relaxed decoder's logic on S. Since the input is a valid codeword, the relaxed decoder is guaranteed to decode correctly. However, the decoder's queries may be non-smooth. For instance, to decode index i, the decoder might always query index i itself. In this case, the sample-based decoder must sample index i to capture any complete query-set. Since this happens only with probability p, obtaining a constant success probability would require $p = \Omega(1)$, yielding only a trivial linear lower bound.

The sample-based decoder. To solve this, [GL21, GGS26] use the following idea: to decode x_i, they first identify the heavy indices, i.e., indices that the total mass of sets that include them (according to $\mu = \mu_i$) is above some threshold m. Let us use $\langle u \rangle = \{S \subseteq [n] \mid u \in S\}$ to denote all the sets that contain u, and $H = H_m(\mu) := \{u \in [n] \mid \mu(\langle u \rangle) \geq m\}$ is the set of heavy indices. Then, instead of trying to sample the indices in H, *guess* the values of $C(x)$ at H. By a standard counting argument, and since each set in $\mathrm{supp}(\mu_i)$ is of size at most q, we have $|H| \leq \frac{q}{m}$.

Specifically, the sample-based decoder of [GL21] and [GGS26] works as follows, for a parameter $m > 0$. First, it samples each bit of its input codeword with probability p, to get a set of sampled indices $W \subseteq [n]$ and a partial view of the codeword $c|_W$. Next, to decode each x_i, it identifies H_m as discussed above. For each assignment $\kappa \in \{0,1\}^H$ (a "guess"), it checks all of the sets $S \in \mathrm{supp}(\mu)$ such that $S \subseteq W \cup H$. Those are the "local views" that the

[1] It is impossible that there exists x, x' with $C(x)|_S = C(x')|_S = y|_S$ and $x_i \neq x'_i$, since this would contradict the 1-sided error of the decoder. If there is no codeword consistent with $y|_S$, then WLOG the decoder outputs $\perp$, since any other symbol can only increase its soundness error.

decoder can simulate. It then feeds all of those local views into the relaxed decoder, using c for the sampled indices W, and κ for the indices in H.

If there is a consensus (i.e., at least one local view is simulated, and all local views lead the decoder to output the same bit), it outputs that bit as the value of x_i.

Analysis. In the correct iteration, where κ is consistent with $C(x)$, all simulated local views are also consistent with the correct codeword, and therefore, by perfect completeness, yield the correct output bit.

In every other iteration, the simulated views are consistent with a corrupted codeword differing from $C(x)$ on at most $|H|$ coordinates. If $|H| \leq \delta n$, soundness guarantees that a large fraction (with respect to μ) of the query sets are "good", meaning that they lead the decoder to output the correct bit or $\perp$.

The sample-based decoder can output an incorrect bit only if, in some iteration, none of the simulated views corresponds to a "good" query set. Thus, denoting the good query sets as $G \subseteq \mathrm{supp}(\mu)$, we seek a threshold m for which

$$\Pr[\exists S \in G \text{ such that } S \subseteq W \cup H] \geq 1 - \exp(-\Omega(|H|)) \,, \tag{1}$$

where the probability is over the choice of W. This failure bound allows us to take a union bound over all $2^{|H|}$ assignments to H, showing that, with constant probability, the sample-based decoder does not output an incorrect bit in any iteration.

Finding a suitable threshold. There is a tradeoff in choosing the threshold m, which determines the size of H: a larger H makes the requirement $S \subseteq W \cup H$ easier to satisfy, but also requires a smaller failure probability.

While [GL21] used a more direct argument based on the notion of t-daisies to find a suitable threshold m, [GGS26] used the notion of *spreadness*. A distribution over sets μ is said to be (m, r)-spread if, for every non-empty set T, the total mass of sets containing T is at most $m/r^{|T|}$. Using Janson's inequality, [GGS26] showed that if a distribution is (m, r)-spread, then Equation 1 holds for $p = \widetilde{\Theta}(1/r)$.

A complementary Lemma of [GGS26] is the *spreadness extraction lemma*, which we also use in this work. Roughly speaking, the lemma states that for every distribution over sets of size at most q, there exists an m such that after removing the heavy coordinates H_m from every query set and restricting to a large subfamily, the resulting distribution is $(m, n^{\Theta(1/q)})$-spread. Together, these two results show that there exists an m for which Equation 1 holds with $p = n^{-\Theta(1/q)}$, yielding the nearly tight lower bound of [GGS26].

2.3 The Challenge of Adaptivity

An adaptive relaxed decoder may choose to query coordinates depending on the answers to its previous queries. This raises two immediate challenges:

- **No single distribution over sets.** The spreadness extraction lemma operates on a fixed distribution over query sets. It is therefore unclear how to apply it, or any other method formulated for such a distribution, in the adaptive setting.
- **No input-independent notion of "heavy" indices.** Since the query distribution depends on the input, so do the degrees of the individual coordinates. Therefore, for a threshold m, it is unclear which set should play the role of H_m.

To overcome these challenges, we exploit the particular structure imposed by adaptivity.

Distributions over decision trees. Any adaptive algorithm can be described by sampling from a distribution over *decision trees*, evaluating the sampled tree on the input y, and outputting the value at the corresponding leaf. Accordingly, any adaptive relaxed decoder with query complexity q can be fully described by a collection of distributions over decision trees, $\{\tau_i\}_{i\in[k]}$, where every tree has depth at most q.[2]

For any fixed index i and input y, each decision tree collapses to a path. Hence, for this y the distribution over decision trees induces a distribution over query sets, denoted $\tau^y = \tau_i^y$. This suggests the following tempting but flawed approach.

A naive solution. Consider $\tau^{C(x)}$, the set distribution corresponding to the operation of the relaxed decoder on the correct codeword. This is a distribution over query sets, as in the non-adaptive case, so we can attempt to apply the spreadness extraction lemma to obtain a threshold m and a set of heavy indices H, and then proceed as in the non-adaptive case.

However, this approach faces two problems:

- **Unknown input codeword.** The sample-based decoder does not know $C(x)$, and hence cannot base its operation on $\tau^{C(x)}$ or on a heavy set extracted from it.
- **Circular dependency.** To enumerate assignments κ to H, the decoder must first determine H, but each assignment induces a different input y, whose query distribution τ^y may have a different set of heavy indices.

Previous approaches. A well-known way to reduce decision trees to sets, which was used in [GL21], is to query all branches of each decision tree. However, doing so can increase the query complexity *exponentially*, from q to 2^q, leading to a much worse lower bound.

Dall'Agnol, Gur, and Lachish [DGL23] extended the $n = k^{1+\Omega\left(\frac{1}{q^2\log^2 q}\right)}$ lower bound of [GL21] to adaptive, 2-sided error RLDCs while avoiding this exponential query blowup. We next describe their approach and the obstacles to obtaining a stronger lower bound within the same framework.

Flattening the tree distribution. The first step in [DGL23] is to replace the distribution over decision trees by a single "flat" multiset containing the query set of every branch of every tree. This yields an **input-independent** structure, allowing notions such as coordinate weights to be defined globally and methods for set distributions to be applied.

The cost is that the degree of a coordinate under the distribution τ^y induced by a particular input y can be larger by a factor of up to 2^q than its degree in the flat distribution. Hence, to capture every coordinate whose weight may exceed m for some input, the global heavy set must include all coordinates of weight at least $m/2^q$ in the flat distribution. This can increase the size of H by a factor of 2^q, causing a corresponding loss in the resulting lower bound.

For constant q, as considered in [DGL23], this loss is absorbed into the constants. In the regime of $q = \omega(1)$, however, this yields a non-constant loss in the lower bound.

[2] This model also captures 2-sided error RLDCs. Two different trees might give the same query set for some specific input, but reach different output leaves.

Lossless partitioning. After the previous step, [DGL23] partitions the flat query distribution into $q+1$ structures, called t-daisies. For every fixed input y, at least one of these daisies has mass $\Omega(1/q)$ under τ^y. To ensure that this daisy still contains sufficiently more mass on correct local views than on erroneous ones, the soundness error must therefore be $O(1/q)$.

Accordingly, as a preprocessing step, [DGL23] performs soundness amplification, reducing the relaxed decoder's soundness error to $O(1/q)$ by repeating it $\Theta(\log q)$ times. This results in a corresponding blowup in query complexity and hence a loss in the final lower bound.

In the non-adaptive setting, techniques developed in [Gol23, GGS26] avoid this loss by extracting a single structure whose mass is arbitrarily close to 1. However, *this does not extend to the adaptive setting*: a small fraction discarded from the global flat distribution may contain all query sets active on some specific input y. Thus, within the framework of [DGL23], the lossless partitioning, the soundness amplification and the resulting $O(\log q)$ query-complexity blowup appear unavoidable.

The limitations of consensus. The above partitioning raises yet another difficulty. Checking consensus separately within each part is insufficient, since all good active paths for a specific input may belong to other parts. On the other hand, checking consensus across all parts simultaneously requires guessing a common assignment to their heavy indices, so that all simulated paths correspond to the same input. The resulting heavy set may be too large, requiring a soundness error probability smaller than what the analysis can guarantee.

[DGL23] addresses this difficulty by replacing consensus with voting. For each assignment, the sampled local views cast votes, and an output is declared only if its number of votes crosses a threshold θ. However, this strategy introduces a new challenge in the soundness analysis.

Upper-tail concentration. As discussed above, for any fixed guess, the probability that the sample-based decoder outputs an incorrect bit must be exponentially small in the number of heavy indices. With the consensus rule, this is done by identifying, for each wrong guess, a family of good sets, and showing that, except with exponentially small probability, at least one of them is contained in $W \cup H$. Equivalently, if X denotes the number of captured good sets, we seek a bound of the form

$$\Pr[X = 0] \leq \exp(-\Omega(|H|)) \ .$$

While [GL21] establishes this bound using a large family of disjoint sets, [GGS26] extracts a spread distribution and applies Janson's inequality, allowing a smaller sampling probability.

With voting, soundness instead requires an upper-tail bound of the form

$$\Pr[Y \geq \theta] \leq \exp(-\Omega(|H|)) \ ,$$

where Y is the number of votes for an incorrect output. To prove this bound, [DGL23] uses the Hajnal–Szemerédi theorem to obtain disjoint sets with independent sampling events, allowing Chernoff bounds for both tails.

One might expect that the spreadness extraction and Janson's inequality used in [GGS26] would yield the required upper-tail bound while retaining the smaller sampling probability. However, **Janson's inequality provides exponential lower-tail concentration but no symmetric upper-tail bound**. This is not merely a limitation of Janson's inequality. In contrast to disjointness, spreadness alone does not imply a corresponding exponential *upper-tail* bound, as we demonstrate next.

Spreadness does not imply exponential upper-tail concentration. We construct an (m, r)-spread distribution for which $\Pr[X \geq 2\,\mathbb{E}[X]]$ is much larger than $\exp(-\Omega(1/m))$, where X is the number of sampled sets for $p = 1/r$. Such a bound is required for soundness, since the heavy set H_m may have size $\Theta(q/m)$.

Consider r vertex-disjoint copies of the clique K_r, and let μ be the uniform distribution over their edges. This is a set family of r^2 base elements and $r\binom{r}{2} = \frac{r^2(r-1)}{2}$ sets, and the mass of each set is $\frac{2}{r^2(r-1)}$. It is straightforward to verify that μ is (m, r)-spread for $m = \frac{2}{r-1}$. The probability of sampling each fixed set is $p^2 = r^{-2}$, and hence $2\,\mathbb{E}[X] = \frac{2}{r^2} \cdot \frac{r^2(r-1)}{2} = r - 1$.

Fix a clique and $2\sqrt{r}$ elements in it, and consider the event that all of these elements are sampled. On this event, $X \geq \binom{2\sqrt{r}}{2} \approx 2r > 2\,\mathbb{E}[X]$. On the other hand, the probability of this event is $\left(\frac{1}{r}\right)^{2\sqrt{r}} \approx \exp(-2\sqrt{r}\log r)$. Hence, $\Pr[X \geq 2\,\mathbb{E}[X]] \geq \exp(-2\sqrt{r}\log r)$, which is much larger than the required $\exp(-1/m) = \exp(-\Theta(r))$.[3]

The Challenge of two-sided error. The upper-tail concentration obstacle also arises for non-adaptive decoders with two-sided error. In this setting, voting is the natural replacement for consensus, which again requires an exponentially small probability that the number of incorrect votes exceeds the threshold. As discussed above, spreadness alone does not provide such a bound.

2.4 Our Approach

We overcome the limitations above using the following ideas:

1. **Guessing the heavy indices:** Instead of finding a global set of heavy coordinates, we enumerate all candidate sets H that could be obtained by applying the spreadness extraction lemma, together with all possible assignments to them.

   Crucially, our soundness analysis does not require spreadness for any of these guesses, and provides a sufficiently strong bound to union bound over all of them.

2. **Decoding via mass estimation:** Instead of using a consensus or voting mechanism, we decode by estimating the probability that the relaxed decoder outputs each bit. We then reduce this task to estimating the total mass of an arbitrary sub-distribution.

3. **The Capped Horvitz–Thompson Estimator:** We develop the CHT estimator, a downward-biased estimator for the total mass of a sub-distribution.

   (a) After normalization by the cap, the CHT estimator is strongly self-bounding, which implies exponential concentration around its mean. This property is sufficient for the soundness analysis.

   (b) We prove that if the estimated sub-distribution is spread, then the bias introduced by capping is small.

   (c) For completeness, we apply the spreadness extraction lemma to the correct-output sub-distribution induced by the correct codeword. The extracted sub-distribution retains sufficient mass and satisfies spreadness, ensuring that the expected CHT estimate is large enough.

[3] We remark that this cannot be fixed by increasing p by a small factor, such as $\log r$. In fact, increasing p from $\frac{1}{r}$ does not remove this obstacle until $p = \Theta(\frac{1}{\sqrt{r}\log r})$.

Our solution exhibits an **asymmetry between soundness and completeness**: we do not need accurate estimates for every candidate. For soundness, the downward bias of CHT is beneficial, and concentration holds regardless of spreadness, allowing us to handle all guesses uniformly even when they induce different query distributions. For completeness, accuracy is required only for the correct guess, where spreadness guarantees that the capping loss is small.

We remark that we use spreadness in a substantially different way from [GGS26]. While [GGS26] uses spreadness to apply Janson's inequality, we use it to directly bound the variance of the discrete derivatives of the Horvitz–Thompson estimator, thereby controlling the bias introduced by capping.

We next describe these ideas in more detail and give a high-level overview of the proof of the main theorem. For simplicity, throughout the overview we track the exponent of the lower bound only up to $1 + \Theta(1/q)$ and do not optimize the constant hidden in the Θ notation.

Our sample-based decoder. Recall that the sample-based decoder receives a valid codeword $c = C(x)$. It samples a set $W \subseteq [n]$ by including each coordinate independently with probability p, and reads $c|_W$. Specifically, we choose

$$p = \frac{q^2 \log n}{n^{\Theta(1/q)}} ,$$

which directly translates to the nearly-tight lower bound.

The sample-based decoder then uses the same sample $(W, c|_W)$ to decode all of x. To decode x_i, it iterates over the $O(q)$ possible thresholds m prescribed by the spreadness extraction lemma, all candidate sets H of the corresponding size $|H| = \lfloor q/m \rfloor$, and all assignments $\kappa \in \{0,1\}^H$.

In each iteration, the decoder considers a hypothetical input string y, obtained from the correct codeword c by replacing its values at the coordinates in H with those specified by κ. We denote the resulting string by $y = c_{H \leftarrow \kappa}$. By our choice of parameters, $|H| \leq \delta n$, where δ is the decoding radius, so y is within the decoding radius of c.

For each output bit $b \in \{0,1\}$, it then estimates the probability that the relaxed decoder outputs b on this string, namely

$$\Pr_{T \sim \tau}[T(y) = b] ,$$

where $T(y)$ denotes the output of the decision tree T on input y, and $\tau = \tau_i$ is the distribution over decision trees corresponding to index i. If the estimate is at least $1/2$, the sample-based decoder outputs b as the value of x_i and halts. Otherwise, it continues to the next guess.

This algorithm raises the obvious question: how can the sample-based decoder estimate $\Pr[T(y) = b]$ if it does not know all of y?

Reduction to mass estimation. The key insight is that for every fixed H, κ, and b, we can decompose this probability according to the set of queried indices outside H.

$$\Pr_{T \sim \tau}[T(y) = b] = \sum_{S \subseteq [n] \setminus H} \Pr_{T \sim \tau}[T(y) = b \text{ and } Q_T(y) \setminus H = S] ,$$

where $Q_T(y) \subseteq [n]$ denotes the coordinates queried by T for the input y. The events in the sum are disjoint, and their union is exactly the event $T(y) = b$.

We define the sub-distribution $\rho_{H,\kappa,b}(S) = \Pr_{T\sim\tau}[T(y) = b \text{ and } Q_T(y) \setminus H = S]$. Then, the probability we seek to estimate is exactly the total mass of this sub-distribution:

$$\Pr_{T\sim\tau}[T(y) = b] = \sum_{S\subseteq[n]\setminus H} \rho_{H,\kappa,b}(S) = \|\rho_{H,\kappa,b}\|_1 \ .$$

This perspective is useful because, for every fixed H and assignment κ (and the induced string y), the sample-based decoder can compute $\rho_{H,\kappa,b}(S)$ for every $S \subseteq W$.

To compute $\rho_{H,\kappa,b}(S)$, the decoder iterates over all trees $T \in \mathrm{supp}(\tau)$ and simulates their execution on y. Queries to indices in H are answered according to κ, and queries to indices in $W \setminus H$ are answered using the sampled view. If, during the simulation, T queries an index outside $H \cup W$, the simulation may be stopped, since in this case $Q_T(y) \setminus H \neq S$. Otherwise, once the simulation is complete, if $T(y) = b$ and $Q_T(y) \setminus H = S$, the decoder adds $\tau(T)$ to a running sum. After iterating over all trees, this sum is exactly $\rho_{H,\kappa,b}(S)$.

The decomposition points us to the following concrete goal:

$$\text{Estimating } \|\rho_{H,\kappa,b}\|_1 \text{ from the values } \rho_{H,\kappa,b}(S) \text{ for sets } S \subseteq W.$$

2.5 The Capped Horvitz–Thompson estimator

As discussed in Section 1.1, a natural tool for the foregoing goal is the Horvitz–Thompson (HT) estimator. Although unbiased, the HT estimator need not concentrate around its mean. To overcome this issue, we introduce the *Capped Horvitz–Thompson* estimator.

We design the CHT estimator so that, after normalization by the cap Γ, it is a *strongly self-bounding* function; concentration inequalities for self-bounding functions due to Boucheron, Lugosi, and Massart [BLM09] (see also [BLM13, Section 6.11]) then imply exponential concentration around its mean.

Roughly speaking, a function f is *strongly* (a, b)*-self-bounding* if (1) for every i and x, its i-th marginal contribution is bounded by 1, namely $0 \le f(x) - f(x^{(i\to 0)}) \le 1$, where $x^{(i\to 0)}$ denotes the vector obtained from x by setting its i-th coordinate to 0; and (2) for every x, the sum of its marginal contributions is bounded in terms of the value of the function itself:

$$\sum_{i\in[n]} (f(x) - f(x^{(i\to 0)})) \le a f(x) + b \ .$$

For a sub-distribution ρ and a fixed x, let $\overline{R}_{\rho,\Gamma}(x) = \sum_S w_S$ where $\{w_S\}$ are optimal weights realizing the maximum in the definition of the Γ-capped HT estimator (CHT; see Definition 1.1). Its marginal contributions satisfy

$$\overline{R}_{\rho,\Gamma}(x) - \overline{R}_{\rho,\Gamma}(x^{(i\to 0)}) \le \sum_{S\ni i} w_S \le \Gamma \ ,$$

where the second inequality holds by the definition of the cap.

Hence, since setting a coordinate to 0 cannot increase the CHT estimate, the normalized CHT estimator $f(x) = \frac{\overline{R}_{\rho,\Gamma}(x)}{\Gamma}$ satisfies the first requirement. The second requirement also holds, with $a = q$ and $b = 0$: by changing the summation order, and since ρ is supported on sets of size at most q,

$$\sum_i (\overline{R}_{\rho,\Gamma}(x) - \overline{R}_{\rho,\Gamma}(x^{(i\to 0)})) \le \sum_i \sum_{S\ni i} w_S = \sum_S |S| w_S \le q \sum_S w_S = q\overline{R}_{\rho,\Gamma}(x) \ .$$

Normalizing by Γ shows that f satisfies the second requirement as well.

For the soundness analysis, we apply the upper-tail concentration inequality for self-bounding functions [BLM13, Theorem 6.19] to f and deduce that if $\mathbb{E}[\overline{R}_{\rho,\Gamma}(X)] \leq \frac{1}{2} - \Omega(1)$, then

$$\Pr[\overline{R}_{\rho,\Gamma}(X) > \frac{1}{2}] = \exp(-\Omega(\frac{1}{q\Gamma})) \ .$$

Remarkably, this inequality does not rely on any structural properties of the estimated sub-distribution.

Similarly, for the completeness analysis, if $\mathbb{E}[\overline{R}_{\rho,\Gamma}(X)] \geq \frac{1}{2} + \Omega(1)$, then the lower-tail concentration inequality for self-bounding functions [BLM13, Theorem 6.20] gives

$$\Pr[\overline{R}_{\rho,\Gamma}(X) \leq \frac{1}{2}] = \exp(-\Omega(\frac{1}{q\Gamma})) \ .$$

Soundness analysis. For brevity, let us denote $\rho = \rho_{H,\kappa,b}$. Let us also denote by $R_\rho, \overline{R}_\rho$ the HT estimator and the capped HT estimator, respectively, for the sub-distribution ρ. Since $\overline{R}_\rho$ is pointwise upper-bounded by the unbiased Horvitz–Thompson estimator, we have

$$\mathbb{E}[\overline{R}_\rho(X)] \leq \mathbb{E}[R_\rho(X)] = \|\rho\|_1 \ .$$

Recall that y is within the decoding radius of the RLDC. Hence, by the soundness guarantee of the relaxed decoder, if b is the incorrect output bit, then

$$\|\rho\|_1 = \Pr_{T\sim\tau}[T(y) = b] \leq 1/3 \ ,$$

and therefore $\mathbb{E}[\overline{R}_\rho(X)] \leq 1/3$.

Thus, by choosing a sufficiently small cap Γ, and using the upper-tail concentration inequality above, we can make the probability of $\overline{R}_\rho(X) > \frac{1}{2}$ small enough to survive a union bound over all guesses. Specifically, we choose $\Gamma = \Theta(\frac{m}{q^2 \log n})$, and obtain an error probability of $\exp(-\Omega(\frac{q \log n}{m}))$ for every fixed guess.

On the other hand, for a fixed m there are $\exp(O(\frac{q \log n}{m}))$ possible guesses of a set H of size $\frac{q}{m}$ and an assignment κ: there are

$$\binom{n}{\lfloor \frac{q}{m} \rfloor} \leq n^{q/m} = O(2^{\frac{q}{m} \log n})$$

possible sets, each with $2^{q/m}$ possible assignments. By choosing the constants appropriately, the concentration bound is strong enough to take a union bound over all these pairs.

Finally, and crucially, there are only $\Theta(q)$ possible values of the threshold m. Taking one additional union bound over these values implies that, with constant probability of, say, 5/6, no iteration of the sample-based decoder outputs the incorrect bit. This completes the soundness analysis.

Spreadness implies small capping loss The key claim for the completeness argument is that if an arbitrary sub-distribution σ is spread, and Γ is chosen appropriately, then $\mathbb{E}[\overline{R}_\sigma(X)]$ is close to $\mathbb{E}[R_\sigma(X)]$. Consider the marginal contributions of R_σ,

$$d_i(x) = \sum_{S \in \langle i \rangle} \sigma(S) p^{-|S|} x_S \ .$$

Like the HT estimator itself, these functions are unbiased: $\mathbb{E}[d_i(X)] = \sigma(\langle i \rangle)$. However, they need not concentrate around their expectations. Their variance may be large, causing $d_i(X)$ to occasionally exceed the cap by a large amount, in which case capping may incur a significant loss.

We show that if σ is spread, then the variance of each d_i, conditioned on i being sampled, is small. Specifically, we prove that if σ is (m, r)-spread, then for $\Gamma = 2m/pr$ the expected total overload (i.e., the sum, over all indices, of the amount by which $d_i(X)$ exceeds the cap) is at most $O\left(\frac{q^2}{pr}\right)$. Thus, if $pr = \omega(q^2)$, this gap can be made arbitrarily small. We then show that the gap $\mathbb{E}[R_\sigma(X)] - \mathbb{E}[\overline{R}_\sigma(X)]$ is bounded by this expected total overload.

Completeness analysis. Our goal is to show there exists an iteration, i.e., a threshold m, a set H, and assignment κ, for which the sample-based decoder outputs the correct bit $b^* = x_i$ with high probability. Denoting $\rho = \rho_{H,\kappa,b^*}$, we want to show $\overline{R}_\rho(X) \geq 1/2$ with probability at least $5/6$.

To this end, we define the sub-distribution

$$\nu(S) = \Pr_{T \sim \tau}[T(C(x)) = b^* \wedge Q_T(C(x)) = S] \,.$$

By the completeness guarantee of the relaxed decoder, $\|\nu\|_1 = \Pr_{T\sim\tau}[T(C(x)) = b^*] \geq 2/3$.

We normalize ν and apply the spreadness extraction lemma [GGS26, Lemma 5.4]. We obtain a threshold m^*, a set H^* with $|H^*| = \lfloor \frac{q}{m^*} \rfloor$, and a sub-distribution $\sigma \leq \nu^{\circ H^*}$ that is (m^*, r)-spread and satisfies $\|\sigma\|_1 \geq 0.99\|\nu\|_1 \geq 0.99 \cdot \frac{2}{3}$.

Next, we consider the sub-distribution used by the sample-based decoder for the correct guess, $\rho = \rho_{H^*,\kappa^*,b^*}$, where $\kappa^* = C(x)|_{H^*}$. By definition,

$$\rho(S) = \Pr_{T \sim \tau}[T(C(x)) = b^* \wedge Q_T(C(x)) \setminus H^* = S] \,.$$

This is exactly the sub-distribution obtained by puncturing ν by H^*. That is, $\rho = \nu^{\circ H^*}$.

We then apply the key claim above to the spread sub-distribution σ, and deduce that $\mathbb{E}[R_\sigma(X)] - \mathbb{E}[\overline{R}_\sigma(X)] = o(1)$, and hence $\mathbb{E}[\overline{R}_\sigma(X)] \geq 0.99 \cdot \frac{2}{3} - o(1)$. Since $\rho = \nu^{\circ H^*} \geq \sigma$, and since the CHT estimator is monotone, we also obtain $\mathbb{E}[\overline{R}_\rho(X)] \geq 0.99 \cdot \frac{2}{3} - o(1)$. In particular, for sufficiently large n, this expectation is bounded away from $1/2$ by a constant.

Finally, we apply the lower-tail concentration inequality for self-bounding functions [BLM13, Theorem 6.20] and deduce that $\Pr[\overline{R}_\rho(X) \leq 1/2]$ is sufficiently small, which completes the completeness proof.

Putting everything together. Combining the soundness and completeness analyses, the sample-based decoder recovers each message bit with probability at least $2/3$ while sampling every coordinate independently with probability $p = \frac{q^2 \log n}{n^{\Theta(1/q)}}$, yielding the stated lower bound.

3 Preliminaries

We provide notation used throughout the paper, together with known facts and theorems.

3.1 Basic notation and facts

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 $x \in U$.

Sub-distributions. Let D be a discrete domain.

- A *sub-distribution* μ over D is a function $\mu : D \to [0,1]$ such that $\sum_{x \in D} \mu(x) \leq 1$. If $\sum_{x \in D} \mu(x) = 1$, we say that μ is a (probability) *distribution*.
- For any subset $A \subseteq D$, the *mass* of A is $\mu(A) = \sum_{x \in A} \mu(x)$.
- The *total mass* of a sub-distribution μ is $\|\mu\|_1 = \sum_{x \in D} \mu(x)$.
- The *support* of a sub-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 sub-distribution μ on a set $A \subseteq D$ with positive mass is a distribution over A, denoted μ_A, defined for each $x \in A$ by $\mu_A(x) = \frac{\mu(x)}{\mu(A)}$.
- The *restriction* of a sub-distribution μ to a set $A \subseteq D$ is a sub-distribution over D, denoted $\mu|_A$, defined for each $x \in D$ by

$$(\mu|_A)(x) = \begin{cases} \mu(x) & x \in A, \\ 0 & x \notin A. \end{cases}$$

  Fact 3.1. *For every sub-distribution μ and a set $A \subseteq D$ with positive mass, $\mu|_A(x) = \mu(A)\mu_A(x)$ for every $x \in A$.*

Additional notations and facts.

- For a value $a \in \mathbb{R}$, we denote $a_+ = \max(a, 0)$.
- We denote $[n] = \{1, 2, 3, \ldots, n\}$.
- All of the logarithms in this paper are in base 2.

Fact 3.2. *For every $b > a > 0$,*

$$\sum_{i=1}^{\infty} \left(\frac{a}{b}\right)^i = \frac{a}{b-a} \ .$$

3.2 Information Theory and Entropy

Definition 3.3. *For $t \in (0,1)$, the binary entropy function is*

$$h_2(t) = -t \log t - (1-t) \log(1-t) \ .$$

Definition 3.4. *The* entropy *of a discrete random variable X is defined by*

$$H(X) = -\sum_x \Pr[X = x] \log \Pr[X = x] \ .$$

Definition 3.5. *For random variables X and Y, the conditional entropy of X given Y is*

$$H(X \mid Y) = \sum_y \Pr[Y = y] H(X \mid Y = y) \ .$$

Fact 3.6 (Fano's inequality for binary random variables [Fan61])**.** *Let $X \to Y \to \widehat{X}$ be a Markov chain, where X and $\widehat{X}$ are binary random variables. Let $\varepsilon = \Pr[\widehat{X} \neq X]$. Then $H(X \mid Y) \leq h_2(\varepsilon)$.*

Fact 3.7 (Chain rule for conditional entropy)**.** *For random variables X, Y, Z,*

$$\begin{aligned} H(X, Y \mid Z) &= H(X \mid Z) + H(Y \mid X, Z) \\ &= H(Y \mid Z) + H(X \mid Y, Z) \ . \end{aligned}$$

In particular, if Y is determined by X and Z, then $H(Y \mid X, Z) = 0$, and

$$H(X \mid Y, Z) = H(X \mid Z) - H(Y \mid Z) \ .$$

3.3 Spreadness

Definition 3.8 ((m,r)-spread distributions)**.** *Let μ be a sub-distribution over $\mathcal{P}(U)$ for a finite universe U. Let $r > 1$ and let $m > 0$. We say that μ is* (m,r)-spread *if for any non-empty set $S \subseteq U$,*

$$\mu(\langle S \rangle) \leq \frac{m}{r^{|S|}} \ .$$

Definition 3.9 (Punctured Distribution)**.** *Let μ be a sub-distribution over $\mathcal{P}(U)$. The* punctured distribution *of μ with respect to a set $H \subseteq U$ is a sub-distribution over $\mathcal{P}(U \setminus H)$, denoted $\mu^{\circ H}$, defined, for every $A \subseteq U \setminus H$, by*

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

Equivalently, puncturing maps each set S to $S \setminus H$ while preserving its weight. Note that

$$\sum_{A \in \mathcal{P}(U \setminus H)} \mu^{\circ H}(A) = \|\mu\|_1 \ .$$

Hence, puncturing preserves the total mass, and in particular, if μ is a distribution, then $\mu^{\circ H}$ is also a distribution.

Lemma 3.10. *(The Spreadness Extraction Lemma, Lemma 5.4 in [GGS26]) Let μ be a distribution over $\mathcal{P}_{\leq q}(U)$, where $|U| = n$. Let $d > q$ be any integer.*

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

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

3.4 Relaxed Locally Decodable Codes

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

Definition 3.11. *(Relaxed locally decodable codes) Let $C : \{0,1\}^k \to \{0,1\}^n$ be an error correcting code. A $(q, \delta, \alpha, \beta)$-relaxed decoder for C is a randomized procedure $\mathcal{B}$ that on an explicit input $i \in [k]$, and oracle access to $w \in \{0,1\}^n$, 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 $\Pr[\mathcal{B}^w(i) = x_i] \geq \alpha$.*
2. *(relaxed local decoding) If there exists $x \in \{0,1\}^k$ such that $\mathrm{dist}(w, C(x)) \leq \delta$, then $\Pr[\mathcal{B}^w(i) \in \{x_i, \perp\}] \geq \beta$.*
3. *For every input i and oracle access to any $w \in \{0,1\}^n$, $\mathcal{B}$ makes at most q queries.*

We say that $\mathcal{B}$ 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, to α as its completeness probability, and to β as its soundness probability.

If $\alpha = 1$, we say that the relaxed decoder has 1-sided error. Unless stated otherwise, we assume throughout the paper that $\alpha = \beta = 2/3$.

4 The Capped Horvitz–Thompson Estimator

In this section we present and prove the main tools we need for constructing the sample-based decoder. In Section 4.1, we define the Horvitz–Thompson (HT) Estimator, and present its basic properties. In Section 4.2 we define the *Capped* Horvitz–Thompson (CHT) Estimator, and compare it to the HT estimator.

We then study the concentration and expectation of the CHT estimator. Specifically, in Section 4.3 we prove that the CHT estimator is *strongly self-bounding*, and in Section 4.4 we use this and the BLM inequalities to prove that the CHT concentrates around its mean. In Section 4.5, we prove that if the underlying sub-distribution is *spread*, then the expectation of the CHT estimator is not much smaller than that of the HT estimator.

4.1 The Horvitz–Thompson Estimator

Our starting point is the problem of estimating the total mass of a sub-distribution μ from a set of coordinates, each sampled independently with probability p. A set S is then fully sampled with probability $p^{|S|}$, so weighting its contribution by $p^{-|S|}$ compensates for the

sampling and yields an unbiased estimator. This motivates the Horvitz–Thompson estimator defined below.

Definition 4.1 (Horvitz–Thompson (HT) Estimator)**.** *Let μ be a sub-distribution over subsets of $[n]$ and let $p \in (0,1]$. That is, for every $S \subseteq [n]$, $\mu(S) \geq 0$ and $\|\mu\|_1 = \sum_S \mu(S) \leq 1$.*

The p-Horvitz–Thompson Estimator *for μ is a function $R_{\mu,p} : \{0,1\}^n \to \mathbb{R}$ defined by*

$$R_{\mu,p}(x) = \sum_{S \subseteq [n]} \mu(S) p^{-|S|} x_S \ ,$$

where $x_S = \prod_{i \in S} x_i$. When it is clear from context, we will suppress p and write $R_\mu(x)$.

A basic fact about the HT estimator, which gives the motivation to use it, is that it is an unbiased estimator of mass $\|\mu\|_1$.

Observation 4.2 (The HT estimator is unbiased)**.** *If $X = (X_1, \ldots, X_n)$ is a vector of independent Bernoulli random variables with probability p, then $\mathbb{E}[X_S] = p^{|S|}$ and the expectation of the Horvitz–Thompson Estimator is*

$$\mathbb{E}[R_\mu(X)] = \sum_{S \subseteq [n]} \mu(S) p^{-|S|} \, \mathbb{E}[X_S] = \sum_{S \subseteq [n]} \mu(S) \ .$$

An important aspect of the HT estimator that we study are its *discrete derivatives*.

Definition 4.3 (Discrete Derivative)**.** *Let $f : \{0,1\}^n \to \mathbb{R}$. For an index $i \in [n]$, define the* discrete derivative $\Delta_i f : \{0,1\}^{n-1} \to \mathbb{R}$ *of f with respect to index i as:*

$$\Delta_i f(x^{(i)}) = f(x^{(i \to 1)}) - f(x^{(i \to 0)}) \ ,$$

where $x^{(i \to b)}$ denotes the vector x with its i-th coordinate fixed to b, and $x^{(i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Claim 4.4 (The discrete derivative of HT estimator)**.** *Let μ be a sub-distribution over subsets of $[n]$. Let R_μ be the p-Horvitz–Thompson Estimator for μ for $p \in (0,1]$. Then for every $i \in [n]$,*

$$\Delta_i R_\mu(x^{(i)}) = \sum_{S \in \langle i \rangle} \mu(S) p^{-|S|} x_{S \setminus \{i\}} \ ,$$

where $\langle i \rangle = \{S \subseteq [n] \mid i \in S\}$.

Note that a similar equality holds for the discrete derivative of any multilinear polynomial.

Proof. We express the HT estimator for μ as

$$R_\mu(x) = \sum_{S \subseteq [n]} \mu(S) p^{-|S|} x_S = \sum_{S \notin \langle i \rangle} \mu(S) p^{-|S|} x_S + x_i \sum_{S \in \langle i \rangle} \mu(S) p^{-|S|} x_{S \setminus \{i\}} \ .$$

Now:

$$R_\mu(x^{(i \to 0)}) = \sum_{S \notin \langle i \rangle} \mu(S) p^{-|S|} x_S + 0 \cdot \sum_{S \in \langle i \rangle} \mu(S) p^{-|S|} x_{S \setminus \{i\}} \ ,$$

and

$$R_\mu(x^{(i \to 1)}) = \sum_{S \notin \langle i \rangle} \mu(S) p^{-|S|} x_S + 1 \cdot \sum_{S \in \langle i \rangle} \mu(S) p^{-|S|} x_{S \setminus \{i\}} \ .$$

Hence:

$$\Delta_i R_\mu(x^{(i)}) = R_\mu(x^{(i \to 1)}) - R_\mu(x^{(i \to 0)}) = \sum_{S \in \langle i \rangle} \mu(S) p^{-|S|} x_{S \setminus \{i\}} \ .$$

□

4.2 The Capped Estimator

While the Horvitz–Thompson estimator is an unbiased estimator, it does not necessarily concentrate around its mean. This motivates capping the Horvitz–Thompson estimator (or more precisely, its discrete derivatives) to ensure concentration.

Definition 4.5 (Capped Horvitz–Thompson (CHT) Estimator)**.** *Let μ be a sub-distribution over subsets of $[n]$. Fix a cap $\Gamma > 0$ and $p \in (0, 1]$.*

The Γ-Capped p-Horvitz–Thompson Estimator *for μ is the function $\overline{R}_\mu : \{0,1\}^n \to \mathbb{R}$ defined by:*

$$\overline{R}_{\mu,p,\Gamma}(x) = \max_{\{w_S\}} \sum_{S \subseteq [n]} w_S \ ,$$

where the maximum is subject to:

$$0 \leq w_S \leq \mu(S) p^{-|S|} x_S \ ,$$

and for every $i \in [n]$:

$$\sum_{S \in \langle i \rangle} w_S \leq \Gamma \ .$$

When it is clear from context, we will suppress p and Γ, and write $\overline{R}_\mu(x)$.

We next state and prove a few immediate properties of the CHT estimator.

Claim 4.6. *Fix $p \in (0, 1]$ and $\Gamma > 0$, and let ρ, ρ' be sub-distributions over subsets of $[n]$. Write $R_\sigma = R_{\sigma,p}$ and $\overline{R}_\sigma = \overline{R}_{\sigma,p,\Gamma}$ for $\sigma \in \{\rho, \rho'\}$. The following hold for every $x, y \in \{0,1\}^n$ and every $i \in [n]$.*

1. *$\overline{R}_\rho(x) \leq R_\rho(x)$.*
2. *For every random variable X over $\{0,1\}^n$, $\mathbb{E}[\overline{R}_\rho(X)] \leq \mathbb{E}[R_\rho(X)]$.*
3. *If $x \leq y$ pointwise, then $\overline{R}_\rho(x) \leq \overline{R}_\rho(y)$.*
4. *$0 \leq \Delta_i \overline{R}_\rho(x^{(i)}) \leq \Gamma$.*
5. *If $\rho \leq \rho'$ pointwise, then $\overline{R}_\rho(x) \leq \overline{R}_{\rho'}(x)$.*

Proof. Let w_S be the weights realizing the maximum $\sum_S w_S = \overline{R}_\rho(x)$. By definition,

$$\overline{R}_\rho(x) = \sum w_S \leq \sum_S \rho(S) p^{-|S|} x_S = R_\rho(x) \ .$$

This proves the first item, and the second one follows by taking expectation on both sides of the inequality.

For the third item, notice that if $x \leq y$, then for every set $x_S \leq y_S$. Hence, the feasible set of weights for x is also feasible for y, as

$$w_S \leq \rho(S) p^{-|S|} x_S \leq \rho(S) p^{-|S|} y_S \ ,$$

and the claim follows.

For the fourth item, let $\{w_S^*\}$ be the weights realizing the maximum $\sum_S w_S^* = \overline{R}_\rho(x^{(i\to 1)})$. Define:

$$w'_S = \begin{cases} 0, & i \in S, \\ w_S^*, & i \notin S. \end{cases}$$

Now, the set of weights $\{w'_S\}$ is a feasible solution for $\overline{R}_\rho(x^{(i\to 0)})$. Hence,

$$\overline{R}_\rho(x^{(i\to 0)}) \geq \sum_S w'_S = \sum_{S \not\in \langle i \rangle} w_S^* \,,$$

and then

$$\begin{aligned} \Delta_i \overline{R}_\rho(x^{(i)}) &= \overline{R}_\rho(x^{(i\to 1)}) - \overline{R}_\rho(x^{(i\to 0)}) \\ &\leq \sum_S w_S^* - \sum_{S \not\in \langle i \rangle} w_S^* \\ &= \sum_{S \in \langle i \rangle} w_S^* \\ &\leq \Gamma \,, \end{aligned}$$

where the last inequality follows from the cap constraint. In addition, $\overline{R}_\rho(x^{(i\to 1)}) \geq \overline{R}_\rho(x^{(i\to 0)})$ due to the monotonicity in x (the previous item), and so $\Delta_i \overline{R}_\rho(x^{(i)}) \geq 0$.

For the last item, observe that the weights $\{w_S\}$ are a feasible solution for $\overline{R}_{\rho'}(x)$, since for every S, $w_S \leq \rho(S) p^{-|S|} x_S \leq \rho'(S) p^{-|S|} x_S$. Therefore, $\overline{R}_{\rho'}(x) \geq \sum_S w_S = \overline{R}_\rho(x)$. □

4.3 The CHT estimator is self-bounding

The main notion we use to prove the concentration inequalities for the CHT estimator is that of *(strong) self-bounding functions* (cf. [BLM13, Section 6.11]).

Definition 4.7 (strong (a,b)-self-bounding)**.** *A nonnegative function $f : \{0,1\}^n \to [0,\infty)$ is called* strongly (a,b)-self-bounding *for $a, b \geq 0$ if there exist functions $f_i : \{0,1\}^{n-1} \to [0,\infty)$ such that for all $i \in [n]$ and $x \in \{0,1\}^n$:*

$$0 \leq f(x) - f_i\left(x^{(i)}\right) \leq 1 \,,$$

and

$$\sum_{i=1}^n \left(f(x) - f_i\left(x^{(i)}\right)\right) \leq a f(x) + b \,.$$

Claim 4.8 (CHT estimator is strongly self-bounding)**.** *Let μ be a sub-distribution supported only on sets of size at most q. Let $\overline{R}_\mu$ be the Γ-Capped p-Horvitz–Thompson estimator for $\Gamma > 0$ and $p \in (0,1]$. Then the function $f = \frac{\overline{R}_\mu}{\Gamma}$ is $(q,0)$-strongly self-bounding.*[4]

Before proving the claim, we make the following observation.

[4]When $\Gamma \leq 1$, $\overline{R}_\mu$ is also $(q,0)$-strongly self-bounding. However, using normalization allows us to use stronger concentration bounds.

Observation 4.9. *For any nonnegative function $f : \{0,1\}^n \to [0,\infty)$ and any $x \in \{0,1\}^n, i \in [n]$,*

$$x_i \Delta_i f(x^{(i)}) = f(x) - f(x^{(i\to 0)}) \,.$$

Proof. By definition,

$$x_i \Delta_i f(x^{(i)}) = x_i \left(f(x^{(i\to 1)}) - f(x^{(i\to 0)}) \right) \,.$$

If $x_i = 0$, then $x = x^{(i\to 0)}$ and $x_i \Delta_i f(x^{(i)}) = f(x) - f(x^{(i\to 0)}) = 0$.
If $x_i = 1$, then $x = x^{(i\to 1)}$ and $x_i \Delta_i f(x^{(i)}) = f(x^{(i\to 1)}) - f(x^{(i\to 0)})$. □

The following two properties immediately prove Claim 4.8.

Claim 4.10. *Let μ be a sub-distribution supported only on sets of size at most q. Let $\overline{R}_\mu$ be the Γ-Capped p-Horvitz–Thompson estimator for $\Gamma > 0, p \in (0,1]$. Then, for any $x \in \{0,1\}^n, i \in [n]$,*

1. $0 \le \Delta_i \overline{R}_\mu(x^{(i)}) \le \Gamma$.
2. $\sum_{i=1}^n x_i \Delta_i \overline{R}_\mu(x^{(i)}) \le q \overline{R}_\mu(x)$.

We remark that the first item indicates that the derivatives of a Γ-capped HT estimator are indeed capped by Γ.

Proof. First, $\Delta_i \overline{R}_\mu(x^{(i)}) \ge 0$ by the monotonicity of $\overline{R}_\mu$ in x (Claim 4.6).

Bounding a single derivative. To see that $\Delta_i \overline{R}_\mu(x^{(i)}) \le \Gamma$, let $\{w^*_S\}$ be the weights realizing the maximum

$$\sum_S w^*_S = \overline{R}_\mu(x^{(i\to 1)}) \,.$$

We next claim that

$$\sum_{S \not\in \langle i \rangle} w^*_S \le \overline{R}_\mu(x^{(i\to 0)}) \,.$$

Define:

$$w'_S = \begin{cases} 0, & i \in S, \\ w^*_S, & i \notin S. \end{cases}$$

The set of weights $\{w'_S\}$ is a feasible solution for $\overline{R}_\mu(x^{(i\to 0)})$: For every set S, we have $0 \le w'_S \le \mu(S) p^{-|S|} (x^{(i\to 0)})_S$, since if $i \in S$ then both the bound and w'_S equal zero, and if $i \notin S$ then $w'_S = w^*_S$, which is bounded since $\{w^*_S\}$ is feasible. In addition, for every j, $\sum_{S \in \langle j \rangle} w'_S \le \sum_{S \in \langle j \rangle} w^*_S \le \Gamma$, so the cap constraints also hold.

Hence,

$$\overline{R}_\mu(x^{(i\to 0)}) \ge \sum_S w'_S = \sum_{S \not\in \langle i \rangle} w^*_S \,.$$

And

$$\begin{aligned}\Delta_i \overline{R}_\mu(x^{(i)}) &= \overline{R}_\mu(x^{(i\to 1)}) - \overline{R}_\mu(x^{(i\to 0)}) \\ &\le \sum_S w_S^* - \sum_{S\not\in\langle i\rangle} w_S^* \\ &= \sum_{S\in\langle i\rangle} w_S^* \\ &\le \Gamma \ .\end{aligned}$$

Bounding the derivatives sum. The proof of the second item follows a similar structure.

Let $\{w_S\}$ be the weights realizing the maximum

$$\sum_S w_S = \overline{R}_\mu(x) \ ,$$

and for every $i \in [n]$ define

$$w_S^{(i)} = \begin{cases} 0, & i \in S, \\ w_S, & i \notin S. \end{cases}$$

Now, for every $i \in [n]$ the set of weights $\{w_S^{(i)}\}$ is also a feasible solution for $\overline{R}_\mu(x^{(i\to 0)})$, and hence

$$\overline{R}_\mu(x^{(i\to 0)}) \ge \sum_S w_S^{(i)} = \sum_{S\not\in\langle i\rangle} w_S \ .$$

We get that

$$\begin{aligned}\sum_{i=1}^n x_i \Delta_i \overline{R}_\mu(x^{(i)}) &= \sum_{i=1}^n \left(\overline{R}_\mu(x) - \overline{R}_\mu(x^{(i\to 0)})\right) \\ &\le \sum_{i=1}^n \left(\sum_S w_S - \sum_{S\not\in\langle i\rangle} w_S\right) \\ &= \sum_{i=1}^n \sum_{S\in\langle i\rangle} w_S\end{aligned}$$

By changing the summation order, and since μ is supported on sets of size at most q,

$$\sum_{i=1}^n x_i \Delta_i \overline{R}_\mu(x^{(i)}) \le \sum_S \sum_{i\in S} w_S = \sum_S |S| w_S \le q \sum_S w_S = q\overline{R}_\mu(x) \ ,$$

as required. □

Proof of Claim 4.8. Let $x \in \{0,1\}^n$, and define $f_i(x^{(i)}) = \frac{\overline{R}_\mu(x^{(i\to 0)})}{\Gamma}$. By the first item of Claim 4.10,

$$\frac{\overline{R}_\mu(x)}{\Gamma} - \frac{\overline{R}_\mu(x^{(i\to 0)})}{\Gamma} = \frac{x_i \Delta_i \overline{R}_\mu(x^{(i)})}{\Gamma} \le 1 \ .$$

Similarly, $\frac{x_i \Delta_i \overline{R}_\mu(x^{(i)})}{\Gamma} \geq 0$.

By the second item of Claim 4.10,

$$\sum_{i=1}^{n} \left(\frac{\overline{R}_\mu(x)}{\Gamma} - \frac{\overline{R}_\mu(x^{(i\to 0)})}{\Gamma} \right) = \sum_{i=1}^{n} \frac{x_i \Delta_i \overline{R}_\mu(x^{(i)})}{\Gamma} \leq q f(x) \ .$$

□

4.4 Concentration of the CHT estimator

We now rely on the machinery of self-bounding functions to prove the concentration of the capped Horvitz–Thompson estimator around its mean.

Theorem 4.11 ([BLM13, Theorem 6.19])**.** *Let $X = (X_1, \ldots, X_n)$ be a vector of independent random variables, each taking values in $\{0,1\}$. Let $a > 0$, and let $f : \{0,1\}^n \to [0,\infty)$ be a $(a,0)$-strongly self-bounding function.*[5] *Let $Z = f(X)$. For all $t > 0$:*

$$\Pr[Z \geq \mathbb{E}[Z] + t] \leq \exp\left(-\frac{t^2}{2a\,\mathbb{E}[Z] + at} \right) \ .$$

Theorem 4.12 ([BLM13, Special case of Theorem 6.20])**.** *Let $X = (X_1, \ldots, X_n)$ be a vector of independent random variables, each taking values in $\{0,1\}$.*

Let $a \geq 1/3$, and let $f : \{0,1\}^n \to [0,\infty)$ be a $(a,0)$-strongly self-bounding function.[6] *Let $Z = f(X)$. For $0 < t \leq \mathbb{E}[Z]$,*

$$\Pr[Z \leq \mathbb{E}[Z] - t] \leq \exp\left(-\frac{t^2}{2a\,\mathbb{E}[Z]} \right) \ .$$

Corollary 4.13 (Concentration of the CHT estimator)**.** *Let $X = (X_1, \ldots, X_n)$ be a vector of independent Bernoulli random variables with probability p.*

Let μ be a sub-distribution supported only on sets of size at most q, let $\theta \in (0,1)$, and let $\gamma > 0$. Let $\overline{R}_\mu$ be the Γ-Capped p-Horvitz–Thompson estimator for μ, and denote $Z = \overline{R}_\mu(X)$.

If $\mathbb{E}[Z] \leq \theta - \gamma$, then

$$\Pr[Z \geq \theta] \leq \exp\left(-\frac{\gamma^2}{2q\Gamma} \right) \ .$$

If $\mathbb{E}[Z] \geq \theta + \gamma$, then

$$\Pr[Z \leq \theta] \leq \exp\left(-\frac{\gamma^2}{2q\Gamma} \right) \ .$$

Proof. By Claim 4.8, $f = \frac{\overline{R}_\mu}{\Gamma}$ is a $(q,0)$-strongly self-bounding function. Let

$$Y = \frac{Z}{\Gamma} = f(X),$$

and

$$t = \frac{\gamma}{\Gamma} \ .$$

[5] The theorem is stated in [BLM13] under slightly more general assumptions, which are immediately satisfied by any $(a,0)$-strongly self-bounding function.

[6] Again, the theorem is stated under more general assumptions, which are satisfied by any $(a,0)$-strongly self-bounding function.

Upper tail. Suppose $\mathbb{E}[Z] \le \theta - \gamma$, and thus $\frac{\theta}{\Gamma} \ge \mathbb{E}[Y] + \frac{\gamma}{\Gamma}$.

Applying Theorem 4.11 with $t > 0$ and $a = q$, we get

$$\begin{aligned}
\Pr[Z \ge \theta] &= \Pr[Y \ge \frac{\theta}{\Gamma}] \\
&\le \Pr[Y \ge \mathbb{E}[Y] + \frac{\gamma}{\Gamma}] \\
&\le \exp\left(-\frac{(\gamma/\Gamma)^2}{2q\,\mathbb{E}[Y] + q\gamma/\Gamma}\right) \\
&= \exp\left(-\frac{\gamma^2}{q\Gamma(2\Gamma\,\mathbb{E}[Y] + \gamma)}\right) \\
&\le \exp\left(-\frac{\gamma^2}{2q\Gamma}\right) ,
\end{aligned}$$

where the last inequality follows from $2\Gamma\,\mathbb{E}[Y] + \gamma = 2\,\mathbb{E}[Z] + \gamma \le 2(\theta - \gamma) + \gamma < 2$, as $\theta < 1$ and $\gamma > 0$.

Lower tail. Suppose $\mathbb{E}[Z] \ge \theta + \gamma$, and thus $\mathbb{E}[Y] \ge \frac{\theta+\gamma}{\Gamma} \ge \frac{\gamma}{\Gamma} = t$. We have $q \ge 1 > 1/3$, and $t \le \mathbb{E}[Y]$, so the hypothesis of Theorem 4.12 is satisfied with t and $a = q$. We get

$$\begin{aligned}
\Pr[Z \le \theta] &= \Pr[Y \le \frac{\theta}{\Gamma}] \\
&\le \Pr[Y \le \mathbb{E}[Y] - \frac{\gamma}{\Gamma}] \\
&\le \exp\left(-\frac{(\gamma/\Gamma)^2}{2q\,\mathbb{E}[Y]}\right) \\
&\le \exp\left(-\frac{\gamma^2}{2q\Gamma}\right) ,
\end{aligned}$$

where the last inequality follows from $\mathbb{E}[Y] = \mathbb{E}[Z]/\Gamma \le 1/\Gamma$. □

4.5 Spreadness implies $\mathbb{E}[R]$ is almost $\mathbb{E}[\overline{R}]$

We have shown that the *capped* HT estimator concentrates around its mean. It is left to show that so long as the underlying sub-distribution is *spread*, the expectation of the capped HT estimator is not much smaller than that of the (non-capped) HT estimator, which is the quantity we would like to estimate.

We need the following elementary inequality.

Lemma 4.14. *Let X be a random variable with finite variance, and let $h > 0$. Suppose that $\mathbb{E}[X] \le h$. Then,*

$$\mathbb{E}[(X - 2h)_+] \le \frac{\mathrm{Var}(X)}{4h} .$$

Proof. Let $h^* := \mathbb{E}[X]$. We first argue that for every x,

$$(x - 2h)_+ = \max(x - 2h, 0) \le \frac{(x - h^*)^2}{4h} .$$

If $x \le 2h$, then the left hand side of the inequality is zero, and the inequality holds. Otherwise, denote $x = h(1 + t)$ for $t > 1$. We now have $x - 2h = h(t - 1)$, and $x - h^* \ge x - h = ht$.

Therefore, it is sufficient to show that:

$$h(t-1) \leq \frac{(ht)^2}{4h} \iff 4t-4 \leq t^2 \iff t^2-4t+4 \geq 0 \ .$$

This is equivalent to $(t-2)^2 \geq 0$, which is true.

Next, we take expectation on both sides of the inequality, and get

$$\mathbb{E}[(X-2h)_+] \leq \frac{\mathbb{E}[(X-h^*)^2]}{4h} = \frac{\mathrm{Var}(X)}{4h} \ .$$

□

Lemma 4.15. *Let μ be an (m,r)-spread sub-distribution over $\mathcal{P}([n])$ supported on sets of size at most q. Let $p \in (0,1]$, let R_μ be the p-Horvitz–Thompson estimator for μ, and let $\overline{R}_\mu$ be the Γ-Capped p-Horvitz–Thompson estimator for μ with $\Gamma = 2m/pr$. Further assume that $pr \geq 2(q-1)$.*

Let $X = (X_1, \ldots, X_n)$ be a vector of independent Bernoulli random variables with probability p. Then,

$$\mathbb{E}[R_\mu(X)] - \mathbb{E}[\overline{R}_\mu(X)] = O\left(\frac{q^2}{pr}\right) \ .$$

Proof. For simplicity of notation, we remove the μ subscripts.

We analyze the random variables $\Delta_i R(X^{(i)})$, and relate them to the expectation gap. Specifically, we prove the following three claims:

Claim 4.16. *For every $x \in \{0,1\}^n$,*

$$R(x) - \overline{R}(x) \leq \sum_{i \in [n]} \left(x_i \Delta_i R(x^{(i)}) - \Gamma\right)_+ .$$

Claim 4.17. *For every $i \in [n]$,*

$$\mathbb{E}[\Delta_i R(X^{(i)})] \leq \frac{m}{pr} = \frac{\Gamma}{2} \ .$$

Claim 4.18. *For every $i \in [n]$,*

$$\mathrm{Var}(\Delta_i R(X^{(i)})) \leq \frac{2m(q-1)}{p^3 r^2} \cdot \mu(\langle i \rangle) \ .$$

Before proving the claims, let us show that combined with Lemma 4.14 they imply the Lemma.

First, taking expectation on the inequality of the first claim, we get

$$\mathbb{E}[R(X)] - \mathbb{E}[\overline{R}(X)] \leq \sum_{i \in [n]} \mathbb{E}\left[\left(X_i \Delta_i R(X^{(i)}) - \Gamma\right)_+\right] \ .$$

Now, by Lemma 4.14, and as $\mathbb{E}[\Delta_i R(X^{(i)})] \leq \frac{\Gamma}{2}$, we get

$$\mathbb{E}[(\Delta_i R(X^{(i)}) - \Gamma)_+] \leq \frac{\mathrm{Var}(\Delta_i R(X^{(i)}))}{4 \cdot \Gamma/2} = \frac{pr}{4m} \mathrm{Var}(\Delta_i R(X^{(i)}))$$

Now, since X_i is independent of $\Delta_i R(X^{(i)})$, we have

$$\mathbb{E}[(X_i \Delta_i R(X^{(i)}) - \Gamma)_+] = p\,\mathbb{E}[(\Delta_i R(X^{(i)}) - \Gamma)_+] \leq \frac{p^2 r}{4m} \operatorname{Var}(\Delta_i R(X^{(i)})) \ .$$

Summing over all $i \in [n]$, and using the bound on $\operatorname{Var}(\Delta_i R(X^{(i)}))$, we deduce:

$$\begin{aligned}
\mathbb{E}[R(X)] - \mathbb{E}[\overline{R}(X)] &\leq \sum_{i \in [n]} \mathbb{E}\left[\left(X_i \Delta_i R(X^{(i)}) - \Gamma\right)_+\right] \\
&\leq \sum_{i \in [n]} \frac{p^2 r}{4m} \operatorname{Var}(\Delta_i R(X^{(i)})) \\
&\leq \sum_{i \in [n]} \frac{p^2 r}{4m} \cdot \frac{2m(q-1)}{p^3 r^2} \cdot \mu(\langle i \rangle) \\
&= \frac{q-1}{2pr} \sum_{i \in [n]} \mu(\langle i \rangle) \ .
\end{aligned}$$

Opening the definition of $\mu(\langle i \rangle)$, and since μ is supported on sets of size at most q, we get:

$$\sum_{i \in [n]} \mu(\langle i \rangle) = \sum_{i \in [n]} \sum_{S \in \langle i \rangle} \mu(S) = \sum_S |S| \mu(S) \leq q \sum_S \mu(S) \leq q \ .$$

We conclude that

$$\mathbb{E}[R(X)] - \mathbb{E}[\overline{R}(X)] \leq \frac{q-1}{2pr} \cdot q = O\left(\frac{q^2}{pr}\right) \ .$$

We now return to proving the stated claims.

Proof of Claim 4.16. To bound $R(x) - \overline{R}(x)$, we construct a set of weights $\{w_S\}$, such that:

1. For every S, $0 \leq w_S \leq \mu(S) p^{-|S|} x_S$.
2. For every $i \in [n]$, $\sum_{S \in \langle i \rangle} w_S \leq \Gamma$.
3. $R(x) - \sum_S w_S \leq \sum_{i \in [n]} \left(x_i \Delta_i R(x^{(i)}) - \Gamma\right)_+$.

We construct the weights by "trimming" the weights of the Horvitz–Thompson estimator $R(x)$. Start with $w_S = \mu(S) p^{-|S|} x_S$. Now, iterate over each $i \in [n]$. If $\sum_{S \in \langle i \rangle} w_S \leq \Gamma$ is already satisfied, continue to the next index. Otherwise, reduce the weights $\{w_S\}_{S \in \langle i \rangle}$ by a total amount of

$$\sum_{S \in \langle i \rangle} w_S - \Gamma \ ,$$

distributing the reduction arbitrarily among these weights while keeping them non-negative.

The inequality $w_S \leq \mu(S) p^{-|S|} x_S$ is satisfied for all S, since we only reduce the weights. Similarly, $\sum_{S \in \langle i \rangle} w_S \leq \Gamma$ is satisfied for every i, since we only reduce weights - reductions while processing some index can only "help" other indices.

Now, throughout the construction, we have $\sum_{S \in \langle i \rangle} w_S \leq \sum_{S \in \langle i \rangle} \mu(S) p^{-|S|} x_S = x_i \Delta_i R(x^{(i)})$. While processing index i, we reduce $\sum_S w_S$ by $\left(\sum_{S \in \langle i \rangle} w_S - \Gamma\right)_+ \leq \left(x_i \Delta_i R(x^{(i)}) - \Gamma\right)_+$, hence (3) is also satisfied.

From (1) and (2) we conclude that $\{w_S\}$ are a feasible solution to the maximization problem defining $\overline{R}(x)$, and hence $\overline{R}(x) \geq \sum_S w_S$. This inequality combined with (3) completes the argument. □

Proof of Claim 4.17. By a direct computation,

$$\begin{aligned}
\mathbb{E}[\Delta_i R(X^{(i)})] &= \mathbb{E}[\sum_{S \in \langle i \rangle} \mu(S) \cdot p^{-|S|} X_{S \setminus \{i\}}] \\
&= \sum_{S \in \langle i \rangle} \mu(S) \cdot p^{-|S|} \, \mathbb{E}[X_{S \setminus \{i\}}] \\
&= \frac{1}{p} \sum_{S \in \langle i \rangle} \mu(S) \\
&= \frac{\mu(\langle i \rangle)}{p} \, .
\end{aligned}$$

Since μ is (m, r)-spread, we have $\mu(\langle i \rangle) \leq \frac{m}{r}$, and hence

$$\mathbb{E}[\Delta_i R(X^{(i)})] \leq \frac{m}{pr} = \frac{\Gamma}{2} \, .$$

□

Proof of Claim 4.18. We next bound the variance of $\Delta_i R(X^{(i)})$. First,

$$\begin{aligned}
\mathbb{E}[\Big(\Delta_i R(X^{(i)})\Big)^2] &= \mathbb{E}[\sum_{S,T \in \langle i \rangle} \mu(S)\mu(T) p^{-|S|-|T|} X_{(S \cup T) \setminus \{i\}}] \\
&= \sum_{S,T \in \langle i \rangle} \mu(S)\mu(T) p^{-|S|-|T|} p^{|(S \cup T) \setminus \{i\}|} \\
&= \frac{1}{p^2} \sum_{S,T \in \langle i \rangle} \mu(S)\mu(T) p^{-|(S \cap T) \setminus \{i\}|} \, .
\end{aligned}$$

And since $\mathbb{E}[\Delta_i R(X^{(i)})]^2 = \frac{\mu(\langle i \rangle)^2}{p^2}$ we get

$$\mathrm{Var}(\Delta_i R(X^{(i)})) = \frac{1}{p^2} \sum_{S,T \in \langle i \rangle} \mu(S)\mu(T)(p^{-|(S \cap T) \setminus \{i\}|} - 1) \, .$$

We now iterate over all the pairs of sets $S, T \in \langle i \rangle$ as follows: first, we iterate over every $S \in \langle i \rangle$. Second, over every intersection size $t \in [q]$. Next, over the possible intersection sets R (that is, subsets of S) of size t that contain i. Finally, over every $T \in \langle i \rangle$ such that $S \cap T = R$. The last iteration is where we leverage the spreadness, as the total mass of sets T that include a subset R is bounded.

We get that:

$$\begin{aligned}\sum_{S,T\in\langle i\rangle}\mu(S)\mu(T)(p^{-|(S\cap T)\setminus\{i\}|}-1) &= \sum_{S\in\langle i\rangle}\sum_{t=1}^{q}\sum_{\substack{R\in\binom{S}{t}\\ i\in R}}\sum_{T:S\cap T=R}\mu(S)\mu(T)(p^{-|(S\cap T)\setminus\{i\}|}-1)\\ &= \sum_{S\in\langle i\rangle}\mu(S)\sum_{t=1}^{q}\sum_{\substack{R\in\binom{S}{t}\\ i\in R}}\sum_{T:S\cap T=R}\mu(T)(p^{-t+1}-1)\\ &= \sum_{S\in\langle i\rangle}\mu(S)\sum_{t=1}^{q}(p^{-t+1}-1)\sum_{\substack{R\in\binom{S}{t}\\ i\in R}}\sum_{T:S\cap T=R}\mu(T)\ .\end{aligned}$$

Now, $S\cap T=R \implies T\supseteq R$, and so

$$\sum_{T:S\cap T=R}\mu(T)\le\sum_{T\supseteq R}\mu(T)=\mu(\langle R\rangle)\ ,$$

where the last equality is by definition.

We now use the spreadness assumption: if $|R|=t$, then $\mu(\langle R\rangle)\le\frac{m}{r^t}$. Hence, and as $|S|\le q$,

$$\sum_{\substack{R\in\binom{S}{t}\\ i\in R}}\mu(\langle R\rangle)\le\sum_{\substack{R\in\binom{S}{t}\\ i\in R}}\frac{m}{r^t}=\frac{m}{r^t}\binom{|S|-1}{t-1}\le\frac{m}{r^t}(|S|-1)^{t-1}\le\frac{m}{r^t}(q-1)^{t-1}\ .$$

Next, we calculate the remaining sum over t.

We can skip the first term, with $t=1$, since $p^{-t+1}-1=1-1=0$. In addition, for $t>1$ we have $p^{-t+1}-1\le p^{-t+1}$. Combining this with the previous calculations, we deduce

$$\sum_{t=1}^{q}(p^{-t+1}-1)\frac{m}{r^t}(q-1)^{t-1}\le\frac{m}{r}\sum_{t=2}^{q}\left(\frac{q-1}{pr}\right)^{t-1}\le\frac{m}{r}\sum_{t=2}^{\infty}\left(\frac{q-1}{pr}\right)^{t-1}\ .$$

By the formula for geometric sums (Fact 3.2),

$$\sum_{t=2}^{\infty}\left(\frac{q-1}{pr}\right)^{t-1}=\sum_{t=1}^{\infty}\left(\frac{q-1}{pr}\right)^{t}=\frac{q-1}{pr-q+1}\le\frac{2(q-1)}{pr}\ ,$$

where the last inequality is by the assumption $pr\ge 2(q-1)$.

Finishing the calculations, and since $\sum_{S\in\langle i\rangle}\mu(S)=\mu(\langle i\rangle)$, we get:

$$\begin{aligned}\mathrm{Var}(\Delta_i R(X^{(i)})) &= \frac{1}{p^2}\sum_{S,T\in\langle i\rangle}\mu(S)\mu(T)(p^{-|(S\cap T)\setminus\{i\}|}-1)\\ &\le \frac{1}{p^2}\sum_{S\in\langle i\rangle}\mu(S)\cdot\frac{m}{r}\cdot\frac{2(q-1)}{pr}\\ &= \frac{1}{p^2}\cdot\mu(\langle i\rangle)\cdot\frac{2m(q-1)}{pr^2}\\ &= \frac{2m(q-1)}{p^3r^2}\mu(\langle i\rangle)\ .\end{aligned}$$

□

□

5 Proof of the Lower Bound

In this section, we prove our main theorem.

Theorem 5.1. *Suppose $C : \{0,1\}^k \to \{0,1\}^n$ is a $(q, \delta, \alpha, \beta)$-RLDC where δ, α, β are constants. Suppose that $\alpha + \beta > 1$, and denote $\lambda = \frac{\alpha}{\alpha+\beta-1}$. Then, for every constant $\varepsilon > 0$,*

$$n = \Omega_{\varepsilon,\alpha,\beta}\left(q^{-2}\left(\frac{k}{\log k}\right)^{1+\frac{1}{(\lambda+\varepsilon)q}}\right) .$$

Theorem 1 is a direct corollary of Theorem 5.1, with $\alpha = \beta = 2/3$.

The requirement $\alpha + \beta > 1$ is necessary for a superlinear lower bound. For any α and β satisfying $\alpha + \beta \leq 1$, there exists a trivial 1-query RLDC, with $n = k$ and an arbitrarily large decoding radius. Specifically, consider the identity code $C(x) = x$ and the relaxed decoder that operates as follows: to decode x_i from an input y, it outputs y_i with probability α and $\perp$ with probability $1 - \alpha$. Completeness holds with probability α, and soundness holds, for any y, with probability at least $1 - \alpha \geq \beta$.[7]

The proof of Theorem 5.1 has two parts. First, we define the notion of a *sample-based decoder* of a code, and prove that having such a decoder implies a lower bound on the block-length of the code. Second, we show how to construct a sample-based decoder for any RLDC.

5.1 RLDC Lower Bounds from Sample-Based Decoders

A main idea in the proof is that of sample-based decoders, which we define next.

Definition 5.2 (Sample-Based Decoder)**.** *Let $C : \{0,1\}^k \to \{0,1\}^n$. A* sample-based decoder *with sampling probability $p \in (0,1]$ and error probability $\varepsilon \in (0, 1/2)$ for C is a randomized algorithm that operates as follows:*

Given a codeword $c = C(x)$, the algorithm samples a set $W \subseteq [n]$ by including each index independently with probability p, reads $c|_W$, and outputs $\hat{x} \in \{0,1\}^k$, which is a deterministic function of W and $c|_W$. A sample-based decoder must satisfy the following requirement:

For every $x \in \{0,1\}^k$ and every $i \in [k]$,

$$\Pr[\hat{x}_i \neq x_i] \leq \varepsilon ,$$

where the probability is over the choice of W.

Having a sample-based decoder implies a lower bound on the block-length of the code.

Lemma 5.3. *Let $C : \{0,1\}^k \to \{0,1\}^n$. Suppose that C has a sample-based decoder with sampling probability p and error probability ε. Then*

$$pn \geq (1 - h_2(\varepsilon))k ,$$

where h_2 denotes the binary entropy function. In particular, if $\varepsilon < 1/2$ is constant, then $pn = \Omega(k)$.

[7] For an RLDC with non-trivial relative distance, one can instead use any systematic constant-rate code with constant relative distance.

Proof. Let X be uniform over $\{0,1\}^k$, let W be the sampled set, and denote

$$Y = (W, C(X)|_W) \ .$$

For every $i \in [k]$, the decoder predicts X_i from Y with error at most ε. Hence, by Fano's inequality (Fact 3.6),

$$H(X_i \mid Y) \leq h_2(\varepsilon) \ .$$

By subadditivity of conditional entropy,

$$H(X \mid Y) \leq \sum_i H(X_i \mid Y) \leq k h_2(\varepsilon) \ .$$

On the other hand, conditioned on $W = w$, the decoder observes $C(X)|_w$. Since $C(X)|_w$ is a string of length $|w|$, its entropy satisfies $H(C(X)|_w \mid W = w) \leq |w|$. Hence,

$$\begin{aligned} H(X \mid W = w, C(X)|_w) &= H(X \mid W = w) - H(C(X)|_w \mid W = w) \\ &\geq H(X \mid W = w) - |w| \ , \end{aligned}$$

where the equality follows from the chain rule and the fact that $C(X)|_w$ is a deterministic function of X for fixed w (Fact 3.7).

Since W is independent of X, we have $H(X \mid W = w) = H(X) = k$. Averaging over W, we get

$$H(X \mid Y) \geq H(X) - \mathbb{E}[|W|] = k - pn \ ,$$

where $H(X) = k$ since X is chosen uniformly over $\{0,1\}^k$.

Combining the two bounds gives

$$k - pn \leq k h_2(\varepsilon) \ .$$

Therefore,

$$pn \geq (1 - h_2(\varepsilon))k \ .$$

□

5.2 Sample-Based Decoders from RLDCs

The following Lemma states that if a code has a relaxed decoder, then it also has a sample-based decoder.

Lemma 5.4. *Let $C : \{0,1\}^k \to \{0,1\}^n$ be a $(q, \delta, \alpha, \beta)$-RLDC where δ, α, β are constants satisfying $\alpha + \beta > 1$. Let $\varepsilon > 0$ be a fixed constant, let $d = \lfloor (\frac{\alpha}{\alpha+\beta-1} + \varepsilon)q \rfloor + 1$ and let $p = \min(1, D\frac{q^2 \log n}{n^{1/d}})$ for a sufficiently large constant $D = D(\varepsilon, \alpha, \beta)$.*

Then, for sufficiently large n, C has a sample-based decoder with sampling probability p and error probability at most $1/3$.

Remark 5.5. *Notice that if $p = 1$ (that is, $n^{1/d} \leq Dq^2 \log n$), then the sample-based decoder reads* all *of the n-bit codeword, and the lemma is trivial. Hence, without loss of generality, we assume $p < 1$ throughout.*

In the remainder of this subsection we prove Lemma 5.4.

5.2.1 Extracting Spreadness

Before describing the operation of the sample-based decoder, let us set up the main tool on which we are based, the spreadness extraction lemma.

The following is a straightforward adaptation of Lemma 3.10 to sub-distributions.

Corollary 5.6. *(Spreadness Extraction for Sub-distributions) Let μ be a sub-distribution over $\mathcal{P}_{\leq q}(U)$, where $|U| = n$. Let $d > q$ be any integer. Suppose that $\|\mu\|_1 > 0$. Then, there exists $j \in [d]$, a family of sets $\mathcal{D} \subseteq \mathrm{supp}(\mu)$, and a set of elements $H \subseteq U$ such that the sub-distribution $(\mu|_{\mathcal{D}})^{\circ H}$ is $(m, n^{1/d})$-spread, where:*

$$m = q\frac{n^{j/d}}{n}, \qquad |H| = \left\lfloor \frac{n}{n^{j/d}} \right\rfloor = \left\lfloor \frac{q}{m} \right\rfloor, \qquad \mu(\mathcal{D}) \geq \left(1 - \frac{q}{d}\right) \|\mu\|_1 \ ,$$

where $\mu|_{\mathcal{D}}$ is the restriction of μ to $\mathcal{D}$. (It is the sub-distribution that gives weight $\mu(S)$ for every $S \in \mathcal{D}$, and weight zero for every other set. That is, $\mu|_{\mathcal{D}} = \mu(\mathcal{D})\mu_{\mathcal{D}}$).

Proof. Define $\nu = \frac{\mu}{\|\mu\|_1}$. This is a distribution that satisfies the conditions of Lemma 3.10. Hence, there exists $j \in [d]$, $\mathcal{D} \subseteq \mathrm{supp}(\mu)$ and $H' \subseteq U$ such that $(\nu_{\mathcal{D}})^{\circ H'}$ is $(m', n^{1/d})$-spread, where:

$$m' = \frac{q}{\nu(\mathcal{D})} \cdot \frac{n^{j/d}}{n}, \qquad |H'| \leq \frac{n}{n^{j/d}}, \qquad \nu(\mathcal{D}) \geq 1 - \frac{q}{d} \ .$$

Padding. First, we add to H' exactly $\left\lfloor \frac{n}{n^{j/d}} \right\rfloor - |H'|$ arbitrary elements from $U \setminus H'$, and let the resulting set be H. The distribution $(\nu_{\mathcal{D}})^{\circ H}$ is also $(m', n^{1/d})$-spread, since removing universe elements only relaxes the spreadness condition.

Total mass. Now, $\mu = \|\mu\|_1 \nu$, and hence $\mu(\mathcal{D}) = \|\mu\|_1 \nu(\mathcal{D}) \geq \left(1 - \frac{q}{d}\right) \|\mu\|_1$.

Spreadness First, notice that $\nu_{\mathcal{D}} = \mu_{\mathcal{D}}$, since for every $S \in \mathcal{D}$

$$\nu_{\mathcal{D}}(S) = \frac{\nu(S)}{\nu(\mathcal{D})} = \frac{\mu(S)/\|\mu\|_1}{\mu(\mathcal{D})/\|\mu\|_1} = \frac{\mu(S)}{\mu(\mathcal{D})} = \mu_{\mathcal{D}}(S) \ .$$

In addition, by definition $\mu(\mathcal{D}) = \|\mu\|_1 \nu(\mathcal{D})$. Therefore, we have

$$\mu|_{\mathcal{D}} = \mu(\mathcal{D})\nu_{\mathcal{D}} = \|\mu\|_1 \nu(\mathcal{D})\nu_{\mathcal{D}} \ .$$

Hence, for every non-empty $T \subseteq U \setminus H$,

$$\begin{aligned}
((\mu|_{\mathcal{D}})^{\circ H})(\langle T \rangle) &= \|\mu\|_1 \nu(\mathcal{D})((\nu_{\mathcal{D}})^{\circ H})(\langle T \rangle) \\
&\leq \|\mu\|_1 \nu(\mathcal{D}) \frac{m'}{n^{|T|/d}} \\
&= \|\mu\|_1 \frac{qn^{j/d}/n}{n^{|T|/d}} \\
&\leq \frac{qn^{j/d}/n}{n^{|T|/d}} \ .
\end{aligned}$$

Thus, $(\mu|_{\mathcal{D}})^{\circ H}$ is $\left(q\frac{n^{j/d}}{n}, n^{1/d}\right)$-spread. □

5.2.2 Constructing the Sample-Based Decoder

We next describe the operation of the stated sample-based decoder of Lemma 5.4. First, we introduce some definitions.

Definitions. Let T be a decision tree, and $y \in \{0,1\}^n$.

1. We denote by $T(y) \in \{0,1,\perp\}$ the output of the decision tree when y is its input.
2. We denote by $Q_T(y) \subseteq [n]$ the set of indices queried on the active path of T on y.
3. Let $A \subseteq [n]$ and $a \in \{0,1\}^A$.

   We denote by $y_{A \leftarrow a} \in \{0,1\}^n$ the vector defined by $y_{A \leftarrow a}(i) = \begin{cases} a(i) & \text{if } i \in A \\ y_i & \text{if } i \notin A \end{cases}$.
4. We denote the indicator vector of a set $A \subseteq [n]$ by $\mathbf{1}_A \in \{0,1\}^n$.

Construction 5.7. *Recall that $d = \lfloor (\frac{\alpha}{\alpha+\beta-1} + \varepsilon)q \rfloor + 1$. Set $r = n^{1/d}$, and let $M = \{qn^{\frac{j}{d}-1} \mid j \in [d]\}$.*[8] *Note that with these definitions, we have $p \leq D\frac{q^2 \log n}{r}$, and $pr = O(q^2 \log n)$.*

Setting the threshold. *Let us denote*

$$\Lambda = \alpha + \beta - 1, \qquad \lambda = \frac{\alpha}{\Lambda}, \qquad \eta = \frac{\varepsilon\Lambda}{\lambda + \varepsilon} .$$

Since

$$d = \lfloor (\lambda + \varepsilon)q \rfloor + 1 > (\lambda + \varepsilon)q ,$$

the gap between $\alpha(1 - q/d)$ and $1 - \beta$ is greater than η, a constant which depends only on $\varepsilon, \alpha, \beta$:

$$\alpha(1 - q/d) - (1 - \beta) > \frac{\varepsilon\Lambda}{\lambda + \varepsilon} > 0 .$$

We choose the threshold to be the midpoint between these endpoints

$$\theta = \frac{1 - \beta + \alpha(1 - q/d)}{2} .$$

Note that $\alpha(1 - q/d) > \theta + \eta/2$ and $1 - \beta < \theta - \eta/2$.

The sample-based decoder. *The sample-based decoder operates in two phases. First, as it must by its definition, it samples $W \sim \mathrm{Bin}([n], p)$, and queries its input c at indices W.*

In the second phase, for every index $i \in [k]$, it computes deterministically from $(W, c|_W)$ a value $\hat{x}_i \in \{0,1\}$. To describe this phase, and for the rest of the proof, let us fix $i \in [k]$ and let $\tau := \tau_i$ be the distribution of trees that corresponds to the operation of the relaxed decoder for decoding index i. We formally describe the computation of $\hat{x}_i$ from $(W, c|_W)$ in Algorithm 1.

Notice that $\overline{R}_{m,H,\kappa,b}(X)$, as defined in the algorithm, depends on $c_{H \leftarrow \kappa}$, which is unknown to the algorithm. Hence, before we even prove the correctness of the algorithm, we need to show that it can be executed.

[8] Notice that M corresponds to the possible values m can get in Corollary 5.6.

Algorithm 1 Sample-based decoder for the RLDC C for a fixed index $i \in [k]$
Input: A pair $(W, c|_W)$ where $W \subseteq [n]$, $c = C(x)$, and $c|_W$ is the restriction of c to W.
Output: $b \in \{0,1\}$

1. For every $m \in M$, every $H \subseteq [n]$ such that $|H| = \left\lfloor \frac{q}{m} \right\rfloor$, and every $\kappa \in \{0,1\}^H$:
   (a) Let $y = c_{H \leftarrow \kappa}$.
   (b) For every $b \in \{0,1\}$:
      i. Define the sub-distribution $\rho_{H,\kappa,b} : \mathcal{P}([n] \setminus H) \to [0,1]$ by
      $$\rho_{H,\kappa,b}(S) = \Pr_{T \sim \tau}[T(y) = b \ \wedge \ Q_T(y) \setminus H = S] \ .$$
      ii. Let $\overline{R}_{m,H,\kappa,b}$ be the $\frac{2m}{pr}$-Capped p-Horvitz–Thompson estimator for $\rho_{H,\kappa,b}$.
      iii. Denote $X = \mathbf{1}_W$.
      iv. If $\overline{R}_{m,H,\kappa,b}(X) \geq \theta$, halt and output b.
2. If the algorithm did not halt, output 0.

Claim 5.8 (CHT estimator is determined by the sampled view)**.** *For every m, H, κ, b, W, the value $\overline{R}_{m,H,\kappa,b}(X)$ is determined by $c|_W$. That is, the values of c at indices not in W do not affect the CHT estimator.*

Proof. This follows from the definition of the capped Horvitz–Thompson estimator.

For every set $S \not\subseteq W$, we have $X_S = 0$, and therefore the constraints defining the capped Horvitz–Thompson estimator imply $w_S = 0$. Hence, to compute $\overline{R}_{m,H,\kappa,b}(X)$, it suffices to compute $\rho_{H,\kappa,b}(S)$ for sets $S \subseteq W$.

Fix such a set $S \subseteq W$. To determine whether a decision tree T contributes to $\rho_{H,\kappa,b}(S)$, we simulate T on $y = c_{H \leftarrow \kappa}$. Queries in H are answered using κ, and queries in $W \setminus H$ are answered using $c|_W$. If during the simulation T queries an index outside $H \cup W$, then necessarily $Q_T(y) \setminus H \neq S$, and T does not contribute to $\rho_{H,\kappa,b}(S)$. Otherwise, the execution of T can be determined from κ and $c|_W$, and the algorithm determines whether $T(y) = b$ and $Q_T(y) \setminus H = S$. □

5.2.3 Analysis of the Sample-Based Decoder

The following two claims show the correctness of Construction 5.7, and prove Lemma 5.4.

Recall that $b^* = x_i$ is the bit the sample-based decoder needs to output.

Claim 5.9 (Soundness)**.** *With probability at least $5/6$, for every $m \in M, H \in \binom{[n]}{\lfloor q/m \rfloor}$, $\kappa \in \{0,1\}^H$ we have $\overline{R}_{m,H,\kappa,(1-b^*)}(X) < \theta$.*

Claim 5.10 (Completeness)**.** *There exists $m \in M, H \in \binom{[n]}{\lfloor q/m \rfloor}, \kappa \in \{0,1\}^H$ such that, with probability at least $5/6$, $\overline{R}_{m,H,\kappa,b^*}(X) \geq \theta$.*

Proof of Lemma 5.4 from the soundness and completeness claims. By Claim 5.9, with probability at least $5/6$, no iteration of the algorithm outputs the incorrect bit $1-b^*$. By Claim 5.10,

with probability at least 5/6, there exists an iteration for which the algorithm outputs b^* if it reaches that iteration.

Therefore, by the union bound, with probability at least 2/3 both events occur simultaneously. On this event, the algorithm cannot output $1-b^*$, and either it halts earlier with output b^*, or it reaches the iteration promised by Claim 5.10 and outputs b^*.

Hence, with probability at least 2/3 the algorithm outputs $b^* = x_i$, as required. □

Proof of Claim 5.9, the soundness claim . We first argue that for every fixed m, H, κ:

$$\Pr[\overline{R}_{m,H,\kappa,(1-b^*)}(X) \geq \theta] \leq \exp\left(-\eta^2 \frac{pr}{16qm}\right) .$$

We claim that $|H| \leq \delta n$. By the construction, for some $j \in [d]$ we have $|H| = \left\lfloor \frac{n}{n^{j/d}} \right\rfloor \leq n^{1-1/d}$. By Remark 5.5 we may assume without loss of generality that $n^{1/d} > Dq^2 \log n$. Hence, for sufficiently large n we have

$$|H| \leq n^{1-1/d} \leq \frac{n}{Dq^2 \log n} \leq \delta n ,$$

for any constant δ.

Notice that y differs from the codeword $c = C(x)$ only on indices that are in H. Hence, from the bound of the size of H, y is δ-close to c, and by the soundness guarantee of the relaxed decoder,

$$\Pr_{T\sim\tau}[T(y) = 1-b^*] \leq 1-\beta .$$

Now,

$$\sum_{S\subseteq[n]\setminus H} \rho_{H,\kappa,1-b^*}(S) = \sum_{S\subseteq[n]\setminus H} \Pr_{T\sim\tau}[T(y) = 1-b^* \ \wedge \ Q_T(y)\setminus H = S] = \Pr_{T\sim\tau}[T(y) = 1-b^*] .$$

That is, $\rho_{H,\kappa,1-b^*}$ is a sub-distribution with total weight of at most $1-\beta$. Hence, the expected value of the (non-capped) p-Horvitz–Thompson estimator for $\rho_{H,\kappa,1-b^*}$ is at most $1-\beta$. By the choice of θ and η we get

$$\mathbb{E}[R_{m,H,\kappa,1-b^*}(X)] \leq 1-\beta < \theta - \eta/2 .$$

Now, recall that for every sub-distribution μ we have $\mathbb{E}[\overline{R}(X)] \leq \mathbb{E}[R(X)]$. Hence, by Corollary 4.13, and plugging $\Gamma = 2m/pr$ and $\gamma = \eta/2$, this implies that:

$$\Pr[\overline{R}_{m,H,\kappa,1-b^*}(X) \geq \theta] \leq \exp\left(-\frac{(\eta/2)^2}{2q\Gamma}\right) = \exp\left(-\eta^2\frac{pr}{16qm}\right) .$$

We finish the proof by applying the union bound three times. First, fix m and $H \in \binom{[n]}{\lfloor q/m \rfloor}$, and iterate over all $\kappa \in \{0,1\}^H$. There are $2^{|H|} = 2^{\lfloor q/m \rfloor}$ possible assignments κ. Therefore, for every fixed m and H,

$$\Pr[\exists \kappa \in \{0,1\}^H : \overline{R}_{m,H,\kappa,1-b^*}(X) \geq \theta] \leq 2^{q/m} \cdot \exp\left(-\eta^2\frac{pr}{16qm}\right) .$$

Second, fix m, and iterate over all subsets $H \in \binom{[n]}{\lfloor q/m \rfloor}$. There are $\binom{n}{\lfloor q/m \rfloor} \leq n^{q/m} = 2^{\frac{q}{m}\log n}$ such subsets. Hence, for every fixed m:

$$\Pr[\exists H \in \binom{[n]}{\lfloor q/m \rfloor}, \exists \kappa \in \{0,1\}^H : \overline{R}_{m,H,\kappa,1-b^*}(X) \geq \theta] \leq 2^{\frac{q}{m}(1+\log n)} \cdot \exp\left(-\eta^2 \frac{pr}{16qm}\right) .$$

Third, we iterate over all d possible values of m. Using $2^x \leq \exp(x)$, for every $m \in M$ we have

$$2^{\frac{q}{m}(1+\log n)} \cdot \exp\left(-\eta^2 \frac{pr}{16qm}\right) \leq \exp\left(\frac{q(1+\log n) - \eta^2 pr/16q}{m}\right) .$$

Assuming $pr > \frac{16q^2(1+\log n)}{\eta^2}$, the numerator in the exponent is negative. Since $m \leq q$, we conclude that

$$\Pr[\exists m, H, \kappa : \overline{R}_{m,H,\kappa,1-b^*}(X) \geq \theta] \leq d \exp\left(1 + \log n - \eta^2 \frac{pr}{16q^2}\right) .$$

By our choice of parameters,

$$pr = Dq^2 \log n .$$

Therefore,

$$d \exp\left(1 + \log n - \eta^2 \frac{pr}{16q^2}\right) = d \exp\left(1 + \log n - \frac{\eta^2 D}{16} \log n\right) .$$

Since $d = \lfloor (\lambda + \varepsilon)q \rfloor + 1 \leq (\lambda + \varepsilon)n + 1$, for a sufficiently large constant D the above quantity is smaller than $1/6$.

Hence, with probability at least $5/6$, for every $m \in M$, every $H \in \binom{[n]}{\lfloor q/m \rfloor}$, and every $\kappa \in \{0,1\}^H$,

$$\overline{R}_{m,H,\kappa,1-b^*}(X) < \theta ,$$

which proves the claim. □

Proof of Claim 5.10, the completeness claim. We prove the completeness claim by analyzing the relaxed decoder on the valid codeword $c = C(x)$.

The completeness sub-distribution. Define the sub-distribution over sets

$$\nu(S) = \Pr_{T \sim \tau}[T(c) = b^* \wedge Q_T(c) = S] .$$

In words, ν can be described by the following process: Choose at random a decision tree T according to the distribution τ. Then, follow the operation of the tree T on the input c. If its output bit is not b^*, "ignore" this tree. Otherwise, return the set of indices queried by T.

Now, notice that $\|\nu\|_1$ is exactly the completeness probability of the relaxed decoder:

$$\|\nu\|_1 = \sum_S \Pr_{T \sim \tau}[T(c) = b^* \wedge Q_T(c) = S] = \Pr_{T \sim \tau}[T(c) = b^*] .$$

By the completeness guarantee of the relaxed decoder, we therefore have

$$\|\nu\|_1 \geq \alpha .$$

Extracting spreadness. Each decision tree in $\mathrm{supp}(\tau)$ queries at most q indices, and hence ν is supported on sets of size at most q. We apply Corollary 5.6 to ν, with $d = \lfloor(\lambda+\varepsilon)q\rfloor + 1$. We deduce there exists $j \in [d]$, $\mathcal{D}^* \subseteq \mathrm{supp}(\nu)$ and $H^* \subseteq [n]$ such that the sub-distribution $(\nu|_{\mathcal{D}^*})^{\circ H^*}$ is $(m^*, n^{1/d})$-spread, where:

$$m^* = q\frac{n^{j/d}}{n}, \qquad |H^*| = \left\lfloor \frac{n}{n^{j/d}} \right\rfloor = \left\lfloor \frac{q}{m^*} \right\rfloor, \qquad \nu(\mathcal{D}^*) \geq \left(1 - \frac{q}{d}\right) \|\nu\|_1 \,.$$

The correct guess. Now, define $\kappa^* \in \{0,1\}^{H^*}$ by $\kappa^*(j) = c_j$. With this definition, we have $c_{H^* \leftarrow \kappa^*} = c$. Hence, pointwise,

$$\nu^{\circ H^*} = \rho_{H^*,\kappa^*,b^*} \,,$$

and therefore

$$(\nu|_{\mathcal{D}^*})^{\circ H^*} \leq \rho_{H^*,\kappa^*,b^*} \,.$$

We next claim that m^*, H^*, κ^* are "the correct guess" of the sample-based decoder. That is, we argue that the $\frac{2m^*}{pr}$-Capped, p-Horvitz–Thompson estimator $\overline{R}_{m^*,H^*,\kappa^*,b^*}$ for the sub-distribution ρ_{H^*,κ^*,b^*} satisfies:

$$\Pr[\overline{R}_{m^*,H^*,\kappa^*,b^*}(X) \geq \theta] \geq 5/6 \,.$$

From spreadness to a large expected CHT estimate. Denoting $\sigma = (\nu|_{\mathcal{D}^*})^{\circ H^*}$, we have $\|\sigma\|_1 = \nu(\mathcal{D}^*)$. Let R_σ be the p-Horvitz–Thompson estimator for σ.

Since the uncapped Horvitz–Thompson estimator is unbiased, and by the choice of θ and η,

$$\mathbb{E}[R_\sigma] = \nu(\mathcal{D}^*) \geq \alpha\left(1 - \frac{q}{d}\right) > \theta + \eta/2 \,.$$

Now, let $\overline{R}_\sigma$ be the $\frac{2m^*}{pr}$-Capped, p-Horvitz–Thompson estimator for σ. Since σ is $(m^*, n^{1/d})$-spread, and $pr > 2(q-1)$, we get from Lemma 4.15 that

$$\mathbb{E}[R_\sigma] - \mathbb{E}[\overline{R}_\sigma] = O\left(\frac{q^2}{pr}\right) \,.$$

Since we have $pr = \omega(q^2)$, for sufficiently large n we have $\mathbb{E}[\overline{R}_\sigma] \geq \mathbb{E}[R_\sigma] - \frac{\eta}{4} \geq \theta + \frac{\eta}{4}$.

Next, from the monotonicity of the Capped Horvitz–Thompson estimator, and as $\rho_{H^*,\kappa^*,b^*} \geq \sigma$, we also have

$$\mathbb{E}[\overline{R}_{m^*,H^*,\kappa^*,b^*}] \geq \mathbb{E}[\overline{R}_\sigma] \geq \theta + \frac{\eta}{4} \,.$$

From expectation to high probability. Now, by applying Corollary 4.13 with $\gamma = \frac{\eta}{4}$ and $\Gamma = \frac{2m^*}{pr}$, we get that:

$$\begin{aligned}\Pr[\overline{R}_{m^*,H^*,\kappa^*,b^*}(X) < \theta] &\leq \exp\left(-\frac{(\eta/4)^2}{2q\Gamma}\right) \\ &= \exp\left(-\frac{\eta^2}{32q\Gamma}\right) \\ &= \exp\left(-\eta^2\frac{pr}{64qm^*}\right) \\ &\leq \exp\left(-\eta^2\frac{pr}{64q^2}\right) ,\end{aligned}$$

where the last inequality follows from $m^* \leq q$.

Now, since $pr = \omega(q^2)$ and η is a constant independent of n, for sufficiently large n we have

$$\Pr[\overline{R}_{m^*,H^*,\kappa^*,b^*}(X) < \theta] \leq \frac{1}{6} ,$$

which finishes the proof. □

5.3 Completing the Proof

Proof of Theorem 5.1. Let $C : \{0,1\}^k \to \{0,1\}^n$ be an $(q, \delta, \alpha, \beta)$-RLDC where δ, α, β are constants satisfying $\alpha + \beta > 1$. Let $\varepsilon > 0$ be a fixed constant, and denote $\lambda = \frac{\alpha}{\alpha+\beta-1}$.

By Lemma 5.4, C has a sample-based decoder with error probability $1/3$ and sampling probability $p = \min\left(1, \frac{Dq^2 \log n}{n^{1/d}}\right) \leq \frac{Dq^2 \log n}{n^{1/d}}$ for some constant D and $d = \lfloor(\lambda+\varepsilon)q\rfloor + 1$.

By Lemma 5.3, and absorbing the constant D into the Ω notation, this implies:

$$\frac{q^2 \log n}{n^{1/d}} n = \Omega(k) .$$

Now, $d \leq (\lambda+\varepsilon)q + 1$, and so

$$q^2 \log n \cdot n^{1-\frac{1}{(\lambda+\varepsilon)q+1}} = \Omega(k) .$$

We can assume without loss of generality that $n \leq k^2$, since otherwise the lower bound trivially holds. Under this assumption, $\log n \leq 2 \log k$.

$$q^2 \log k \cdot n^{\frac{(\lambda+\varepsilon)q}{(\lambda+\varepsilon)q+1}} = \Omega(k) .$$

By rearranging the terms we get

$$n = \Omega\left(\left(q^{-2}\frac{k}{\log k}\right)^{\frac{(\lambda+\varepsilon)q+1}{(\lambda+\varepsilon)q}}\right) = \Omega\left(\left(q^{-2}\frac{k}{\log k}\right)^{1+\frac{1}{(\lambda+\varepsilon)q}}\right) .$$

Finally, we observe that $\left(q^{-2}\right)^{1/(\lambda+\varepsilon)q} = \Theta(1)$, and hence $\left(q^{-2}\right)^{1+\frac{1}{(\lambda+\varepsilon)q}} = \Theta(q^{-2})$. The theorem follows. □

Acknowledgments

We would like to thank Irit Dinur, Oded Goldreich, and Guy Rothblum for valuable discussions and helpful insights about relaxed LDCs.

References

[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

[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, 2, 15

[BLM09] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. On concentration of self-bounding functions. *Electronic Journal of Probability*, 14:1884–1899, 2009. 10

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities - A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. 10, 11, 12, 18, 21

[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

[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, 3

[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

[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, 3, 6, 7

[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

[Fan61] Robert M. Fano. *Transmission of Information: A Statistical Theory of Communications*. The MIT Press, Cambridge, MA, 1961. 14

[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, 2

[GGS26] Guy Goldberg, Tom Gur, and Sidhant Saraogi. Nearly tight lower bounds for relaxed locally decodable codes via robust daisies. In Aditya Bhaskara and Artur Czumaj, editors, *Proceedings of the 58th Annual ACM Symposium on Theory of Computing, STOC 2026, Salt Lake City, UT, USA, June 22-26, 2026*, pages 2210–2217. ACM, 2026. 1, 3, 4, 5, 7, 9, 12, 15

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

[GMWZ26] Tom Gur, Dor Minzer, Guy Weissenberg, and Kai Zhe Zheng. 3-query rldcs are strictly stronger than 3-query ldcs. *STOC 2026*, 2026. 1

[Gol23] Oded Goldreich. On the lower bound on the length of relaxed locally decodable codes. *Electron. Colloquium Comput. Complex.*, TR23-064, 2023. 3, 7

[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, 3

[Gol24b] Oded Goldreich. On the relaxed LDC of BGHSV: A survey that corrects the record. *Electron. Colloquium Comput. Complex.*, TR24-078, 2024. 1, 2

[GR18] Tom Gur and Ron D Rothblum. Non-interactive proofs of proximity. *Computational Complexity*, 27(1):99–207, 2018. 1, 2

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

[HT52] Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952. 2

[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

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

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il