Probabilistic Variants of Rota’s so-called “Critical Problem” in Combinatorics and Coding Theory

Jason Altschuler
Advisor: Emmanuel Abbe

Abstract

We present two results in the field of source coding, or data compression. First, we analyze a probabilistic variant of the linear boolean classification (LBC) problem, proposed recently by Abbe et al. 2016. They study the worst-case setting, which is to find the fewest number of linear queries required to determine in which of two given symmetric sets an unknown boolean vector lies. Here, we consider a lossy analogue of LBC, which allows for a vanishing probability of classification error, à la Shannon 1948. We show that if the Hamming distance between the closest vectors in the two classes is linear in the dimension $n$, then we can use arbitrarily few linear queries $m_n$ and still attain vanishing probability of error, provided $m_n \to \infty$ eventually as $n \to \infty$. This gives substantial compression gains: we give examples where lossless LBC requires a linear number of queries, but we can use many fewer (e.g. $\log \log \log \log \log n$) if we allow for a vanishing probability of error. When the classes do not have this separation, we give a simple upper bound via a reduction to Shannon’s Source Coding Theorem, and conjecture a lower bound.

Second, we study the subspace-avoiding-set (SAS) problem, which seeks to find the largest linear subspace of $F_2^n$ that does not intersect a given symmetric set. This problem generalizes lossless linear source coding; however, it is known to be an instance of the so-called “critical problem” proposed by Crapo and Rota 1970. Here, we study a probabilistic variant of SAS which allows for a vanishing fraction of intersections (errors). Our main result is a lower bound on how large the subspaces can be as a function of the number of allowed intersections. When no errors are allowed, our result matches the well-known Gilbert-Varshamov bound, which is conjectured to be tight (Goppa’s conjecture). Our result also shows an interesting paradox: the existence of arbitrarily large subspaces $X_n$ of dimension $n + o(n)$ such that a vector drawn uniformly at random from a given symmetric set will not lie in $X_n$, with probability tending to 1 as $n \to \infty$. 

## Contents

1 Introduction ......................................................... 3
   1.1 Worst-Case Linear Source Coding: primal and dual problems ....... 3
   1.2 Linear Boolean Classification: worst-case and probabilistic settings .. 4
   1.3 Subspace-Avoiding-Set Problem: worst-case and probabilistic settings .. 4
   1.4 Contributions .................................................. 5
   1.5 Notation ....................................................... 6
   1.6 Outline of article ............................................. 6

2 Geometry of symmetric sets in $\mathbb{F}_2^n$ .......................... 7
   2.1 Preliminaries .................................................. 7
   2.2 Tight bounds on the size of symmetric sets ........................ 7

3 Results for Probabilistic Linear Boolean Classification ................. 10
   3.1 P-LBC for $\Theta(n)$ separated families: vanishing error probability with number of queries growing arbitrarily slowly ............................................................... 10
   3.2 P-LBC for $o(n)$ tight families: simple upper bound and conjecture on lower bound .............. 12

4 Results for Probabilistic Subspace-Avoiding-Set Problem ............... 14
   4.1 Lower bound on $\gamma$-P-SAS capacities for symmetric sets: statement, corollaries, and intuition 14
   4.2 Proof of lower bound via probabilistic method ................................................................. 15

5 Future work .................................................................. 19

6 Bibliography .................................................................. 20

A Appendix .................................................................. 21
   A.1 Proof of Lemma 5 .................................................. 21
   A.2 Proof of Lemma 6 .................................................. 21
1 Introduction

1.1 Worst-Case Linear Source Coding: primal and dual problems

A fundamental problem in information theory is source coding, also called data compression. Informally, the goal is to map a set of messages into a lower dimensional space in order to remove redundancy. This is well-studied in both the lossless (worst-case) and lossy (probabilistic) regimes. Lossless techniques require exact recovery of messages, whereas lossy ones allow for a small probability of decoding error in exchange for better compression.

More precisely, the problem of lossless source coding is to find the smallest dimension $m$ such that a given set $S \subseteq \mathbb{F}_2^n$ can be mapped injectively into $\mathbb{F}_2^m$. Such a mapping clearly allows for exact recovery of messages. Without any constraints on the mapping, this problem is trivial since $m = \lceil \log_2 |S| \rceil$ suffices by simply indexing $S$. However, such an indexing map requires an exponentially sized look-up table, rendering it computationally inefficient. As such, a typical constraint in coding theory is to require the map to be linear; this ensures both fast compression and de-compression, as well as various other desirable properties we will discuss shortly.

To recap, the problem of lossless linear source coding seeks the lowest-rank linear map $T : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ that is $S$-injective, meaning that

$$x, x' \in S, x \neq x' \implies T(x) \neq T(x')$$

Because we are working over $\mathbb{F}_2$ and $T$ is linear, this is clearly equivalent to $T(x + x') \neq 0$ for all distinct $x, x' \in S$. This is in turn equivalent to $T(x) \neq 0$ for all $x \in S$, where $S_+$ denotes the (Minkowski) sumset $S + S = \{x + x' : x, x' \in S\}$. Thus the $S$-injectivity property can be restated as:

$$\ker(T) \cap (S_+ \setminus \{0^n\}) = \emptyset$$

The equivalence of the conditions in equations (1) and (2) gives us two dual phrasings of lossless linear source coding, which we formalize in the following definition.

Definition 1 (Worst-Case Linear Source Coding problem (WC-LSC)). Given a set $S \subseteq \mathbb{F}_2^n$, we ask:

- **Primal WC-LSC**: Find the smallest rank $m$, such that there exists a linear map $T : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m$ that is $S$-injective.
- **Dual WC-LSC**: Find the dimension of the largest linear subspace of $\mathbb{F}_2^n$ that does not intersect $S_+ \setminus \{0^n\}$.

Intuitively, the primal problem seeks to maximize the rank of $T$, whereas the dual problem aims to minimize its nullity. By the Rank-Nullity Theorem, $m^\ast(T)(S) + m^\ast(T)(S_+) = n$, where $m^\ast(T)(S)$ and $m^\ast(T)(S_+)$ denote the optimal solutions to the primal and dual problems, respectively.

Importantly, Dual WC-LSC is an instance of the so-called “critical problem” posed by Crapo and Rota [1970], and is an open problem for arbitrary sets. Because of the equivalence between Primal WC-LSC and Dual WC-LSC, this means that the primal problem (i.e. lossless source coding) is also hard in general.

Thus, instead of exactly finding the optimal compression dimension $m^\ast(S)$, the literature has focused on finding upper and lower bounds. Despite extensive work in Hamming [1950], Plotkin [1960], Johnson [1962], Singleton [1964], Bassalygo [1965], Jiang and Vardy [2004], Vu and Wu [2005], there is a still a large gap between the best known lower bound (due to McEliece et al. [1977]) and the best known upper bound (due to Varshamov [1967]). Even for Hamming balls, where much attention has been focused, the asymptotically optimal compression dimension $\frac{m^\ast(S)}{n}$ is not even known to converge in the limit as $n \rightarrow \infty$. We refer the reader to van Lint [2012] and Abbe [2014] for further details on the exciting chase for better bounds.

Because lossless source coding is in general “hard” as we discussed, in this article we instead focus on lossy variants. In particular, we study a lossy variant of the primal problem called (probabilistic) linear boolean classification and a lossy variant of the dual problem called the (probabilistic) subspace-avoiding-set problem. These are introduced in the following sections 1.2 and 1.3 respectively.

---

1 We note that this could be done over any finite field $\mathbb{F}_q$, but for simplicity we work in $\mathbb{F}_2$. The results generalize to other fields.
1.2 Linear Boolean Classification: worst-case and probabilistic settings

Recall from Definition 1 that (the primal phrasing of) lossless source coding requires exact decoding of a message \( x \in S \). An interesting relaxation recently proposed by [Abbe et al., 2016] is to require exact decoding of only a characteristic of the message. Specifically, consider compressing a vector \( x \) that lies in one of two disjoint sets \( S_1, S_2 \subseteq \mathbb{F}_2^n \). The goal of linear boolean classification (LBC) is to find the fewest number of linear queries required to classify whether \( x \) is in \( S_1 \) or \( S_2 \). Clearly this is no more than the number of linear queries to completely identify \( x \), which is the goal of Primal WC-LSC.

We formalize this problem of worst-case LBC as follows.

**Definition 2 (Worst-Case Linear Boolean Classification (WC-LBC)).** Given two disjoint sets \( S_1, S_2 \subseteq \mathbb{F}_2^n \), find the smallest rank \( m \) such that there exists a linear map \( T: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m \) that is \((S_1, S_2)\)-distinguishable, i.e. \(Tx_1 \neq Tx_2\) for all \( x_1 \in S_1 \) and \( x_2 \in S_2 \).

[Abbe et al., 2016] find the optimal WC-LBC compression dimension for Hamming annuli under certain conditions, and conjecture that the general case agrees with the Gilbert-Varshamov bound. In this article, however, we will approach LBC from a lossy (probabilistic) standpoint. Informally, this amounts to allowing for an error probability tending to 0 as \( n \rightarrow \infty \) in order to hopefully obtain large compression gains.

As a motivating example, consider each message as a vectorized image \( x \in \mathbb{F}_2^n \). Assume we know beforehand that every image \( x \) either has mostly white pixels (at most \( \alpha n \) ones) or mostly black pixels (at most \( \alpha n \) zeros), for some \( \alpha \in (0, \frac{1}{2}) \). [Abbe et al., 2016] show that lossless LBC requires exactly \( 2[\alpha n] + 1 \) linear queries to errorlessly determine whether a given image was mostly white or mostly black. We will show in this article that, if the sets are separated in a way we make formal later—but is certainly satisfied in this example—then we can take any number of queries that tends to \( \infty \) as \( n \rightarrow \infty \), and still have the probability of classification error tend to 0. For example, we show that \( \log n \) or even \( \log \log \log \log n \) linear queries is sufficient to achieve vanishing probability of error. In application terms, this means large compression gains can be attained by allowing for a vanishing probability of error.

A few technical nuances must be discussed before we can formally state this probabilistic relaxation. First, in order to discuss probability of decoding error, we must take a distribution over the messages. A natural such measure is the uniform mixture model, in which a random message, denoted now by a capital \( X \), is drawn uniformly at random from \( S_1 \) with probability \( \frac{1}{2} \), and otherwise drawn uniformly at random from \( S_2 \). Secondly, in order to make the phrase “error probability tending to 0” make sense, we must take a sequence of sets as \( n \rightarrow \infty \). This is formally stated as follows.

**Definition 3 (Probabilistic Linear Boolean Classification (P-LBC)).** Let \( \mathcal{F}_1 = \{ S_{1,n} \subseteq \mathbb{F}_2^n \}_{n \in \mathbb{N}} \) and \( \mathcal{F}_2 = \{ S_{2,n} \subseteq \mathbb{F}_2^n \}_{n \in \mathbb{N}} \) satisfy \( S_{1,n} \cap S_{2,n} = \emptyset \) for all \( n \in \mathbb{N} \), and let the measures \( \mu_n \) denote the uniform mixture models over the sets \( S_{1,n} \) and \( S_{2,n} \). Find the asymptotically smallest sequence of ranks \( \{ m_n \}_{n \in \mathbb{N}} \) such that there exists a sequence of linear maps \( \{ T_n: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^{m_n} \}_{n \in \mathbb{N}} \) satisfying:

\[
\lim_{n \rightarrow \infty} \mathbb{P}_{X \sim \mu_n} \left\{ i(X) \neq \underset{i \in \{1,2\}}{\arg \max} \mu_n(\{ x' \in S_{i,n} : T_n(x') = T_n(X) \}) \right\} = 0
\]

where \( i(X) \in \{1,2\} \) is the random variable corresponding to whether \( X \) belongs to the set \( S_{1,n} \) or \( S_{2,n} \).

The argmax term in equation 3 is the maximum a posteriori (MAP) estimation of which set \( X \) was in. Intuitively, the LHS of this equation can be informally thought of as the limiting probability that one incorrectly recovers whether \( X \) is in \( S_{1,n} \) or \( S_{2,n} \) from applying MAP to the observation \( T_nX \). By forcing this probability to tend to 0, we ensure that as \( n \rightarrow \infty \), we can almost surely discern which set \( X \) was in with \( m_n \) linear queries.

We note that the \((S_1, S_2)\)-distinguishability condition for WC-LBC in Definition 2 is equivalent to requiring the non-limiting error probability in equation 3 to be exactly 0.

1.3 Subspace-Avoiding-Set Problem: worst-case and probabilistic settings

We introduced LBC above as a relaxation of Primal WC-LSC; here, we introduce an approach to analyze Dual WC-LSC. We can generalize this dual problem from Definition 1 to more general sets \( S \):
Definition 4 (Worst-Case Subspace-Avoiding-Set problem (WC-SAS)). Given a set \( S \subseteq \mathbb{F}_2^n \), find the dimension of the largest linear subspace that does not intersect \( S \).

Importantly, WC-SAS strictly generalizes several problems that we have already seen. First, observe that WC-SAS is strictly more general than Dual WC-LSC, since not every set in \( \mathbb{F}_2^n \) can be written as a sumset minus \( \{0^n\} \). Consider for example \( \{x \in \mathbb{F}_2^n : \# \text{ of } 1 \text{'s in } x \in \{1,4\} \} \). Secondly, WC-LBC(\( S_1, S_2 \)) reduces to WC-SAS(\( S_1 + S_2 \)) since \( T \) is \( (S_1, S_2) \)-injective in Definition 2 if \( \ker(T) \cap (S_1 + S_2) = \emptyset \).

These reductions bring both good and bad news. On the positive side, it implies that a solution to WC-SAS also gives solutions to Dual WC-LSC and WC-LBC. On the other hand, however, they also imply that WC-SAS is a difficult open problem, since Dual WC-LSC (and thus also the so-called “critical problem” from [Crapo and Rota, 1970]) is a special sub-case. Therefore, instead of approaching SAS in a lossless (worst-case) setting, we instead study a lossy (probabilistic) relaxation.

A natural such relaxation is to allow for a small intersection between the subspace and given set. Recalling our duality proof of the Primal and Dual WC-LSC problems in Section 1.1, these intersection points roughly correspond to the number of messages we might incorrectly decode. A typical constraint in coding theory is that the probability of decoding error tends to 0 as \( n \to \infty \); here, the analogue is to require only a vanishing number of intersections as \( n \to \infty \). This requires taking a sequence of sets, which we formalize as follows.

Definition 5 (Probabilistic Subspace-Avoiding-Set problem (P-SAS)). Let \( \gamma \in [0, 1] \). Given a family \( \mathcal{F} = \{S_n \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}} \) of sets to avoid, a constant \( c \in [0, 1] \) is said to be a \( \gamma \)-P-SAS achievable rate for \( \mathcal{F} \) if there exists a sequence of \([cn]\)-dimensional subspaces \( \{X_n \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}} \) such that

\[
\lim_{n \to \infty} \frac{|S_n \cap X_n|}{|S_n|^\gamma} = 0
\]

Find the \( \gamma \)-P-SAS capacity \( k^*_\gamma(\mathcal{F}) := \sup_{c \in [0, 1]} \{c : c \text{ is a } \gamma \text{-P-SAS achievable rate for } \mathcal{F}\} \).

The reason we define P-SAS with the parameter \( \gamma \in [0, 1] \) is that it allows for a smooth interpolation of how large our subspaces can be (which corresponds to how good our compression rate is) in terms of how much decoding error we are allowed. In particular, the setting \( \gamma = 0 \) requires that \( \lim_{n \to \infty} |S_n \cap X_n| \) which is an asymptotic version of the lossless setting in WC-LSC. The other extreme \( \gamma = 1 \) also has a natural interpretation: that only a vanishing fraction of \( S_n \) intersect the subspace \( X_n \). We can express this formally as the following probability measure:

\[
\lim_{n \to \infty} \mathbb{P}_{Y \sim S_n} \{Y \in X_n\} = \lim_{n \to \infty} \frac{|S_n \cap X_n|}{|S_n|} = 0
\]

where \( Y \sim S_n \) denotes that \( Y \) is drawn uniformly at random from the set \( S_n \).

1.4 Contributions

In this paper, we analyze relaxations of Primal WC-LSC (namely linear boolean classification) and Dual WC-LSC (namely the subspace-avoiding-set problem).

Contributions for Linear Boolean Classification (LBC). Our main result is that if \( \mathcal{F}_1 = \{S_{1,n} \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}} \) and \( \mathcal{F}_2 = \{S_{2,n} \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}} \) are families of general symmetric sets satisfying \( \min_{v_1 \in S_{1,n}, v_2 \in S_{2,n}} d(v_1, v_2) = \Theta(n) \), then we can accomplish P-LBC with ranks \( \{m_n\}_{n \in \mathbb{N}} \) that grow to infinity arbitrarily slowly (e.g. \( m_n = \log \log \log \log \log n \)). This allows for substantial compression gains compared to lossless LBC. In particular, we give examples where lossless LBC requires a linear number of queries, but our result shows that if we allow for a vanishing probability of error, we may use any arbitrarily small number \( m_n \) of linear queries provided \( m_n \to \infty \) eventually.

The setting where \( \mathcal{F}_1 \) and \( \mathcal{F}_2 \) are not “separated”, i.e. \( \min_{v_1 \in S_{1,n}, v_2 \in S_{2,n}} d(v_1, v_2) = o(n) \) is not as clear because intuitively the classes are much more easily confused with a small number of linear queries. We give a simple upper bound for P-LBC via a reduction to Shannon’s famous Source Coding Theorem, but suspect that it is not tight. We conjecture, however, that any P-LBC decoding scheme for “non-separated” families requires the number of linear queries \( m_n \) to scale linearly with \( n \).
Contributions for the Subspace-Avoiding-Set Problem (SAS). Our main result is a lower bound of $1 - (1 - \gamma)\sigma(\mathcal{F})$ for the $\gamma$-P-SAS capacity of a family of symmetric sets $\mathcal{F} = \{S_n \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$, where $\sigma(\mathcal{F}) := \lim_{n \to \infty} \max_{x \in S_n} H\left(\frac{w(x)}{n}\right)$ and $H(\cdot)$ is the Shannon entropy of a Bernoulli random variable. Informally, this shows the existence of large subspaces of dimension $n (1 - (1 - \gamma)\sigma(\mathcal{F})) + o(n)$ that mostly avoid the sets in $\mathcal{F}$, up to the fraction of intersections (errors) allowed by $\gamma$-P-SAS. We note that when $\gamma = 0$, this bound matches the well-known Gilbert-Varshamov bound for lossless source coding [Varshamov, 1957], which is conjectured to be tight [Goppa, 1993]. On the other extreme, when $\gamma = 1$, our bounds show an interesting paradox: the existence of arbitrarily large subspaces $X_n$ of dimension $n + o(n)$ that satisfy $\lim_{n \to \infty} P_{Y \sim S_n}\{Y \in X_n\} = 0$. We note that our bound gives a linear relationship between the error parameter $\gamma \in [0, 1]$ and the $\gamma$-P-SAS capacity (which is informally the dimension of the largest “avoiding” subspaces).

1.5 Notation

Our notation is mostly standard and in general follows [Abbe, 2014] and [Grimmett and Stirzaker, 1992] for coding theory and probability theory, respectively. For completeness, we list a few notations that we use commonly throughout the paper. For $n \in \mathbb{N}$, we write $[n]$ to denote the set $\{1, \ldots, n\}$. We write $\mathbb{F}_2$ to denote the Galois Field of size 2. The **Hamming weight** of a vector $x \in \mathbb{F}_2^n$ is the number of 1’s in it, i.e. $w(x) = \sum_{i=1}^{n} x_i$, where the addition is done over $\mathbb{Z}$ instead of $\mathbb{F}_2$. The **Hamming distance** between two vectors $x, y \in \mathbb{F}_2^n$ is the weight of their sum, i.e. $d(x, y) = w(x + y)$.

We say that a family of random variables $\{X_\alpha\}$ is **independently and identically distributed (i.i.d.)** if the random variables $X_\alpha$ are independently drawn from the same distribution. We say that an indexed family $\{E_n\}_{n \in \mathbb{N}}$ of measurable events occurs **with high probability (w.h.p.)** if $\lim_{n \to \infty} P\{E_n\} = 1$, and with **vanishing probability** if $\lim_{n \to \infty} P\{E_n\} = 0$. Let $S$ and $S'$ be any finite sets, e.g. $S, S' \subseteq \mathbb{F}_2^n$. We write $X \sim S$ to denote that $X$ is drawn uniformly at random from $S$. $X \sim_p (S, S')$ denotes the mixture model where $X$ is drawn uniformly at random from $S$ with probability $p$, and otherwise drawn uniformly at random from $S'$. In particular, we write $X \sim (S, S')$ as shorthand for the uniform mixture model $X \sim \frac{1}{2} (S, S')$. We write $X \sim \text{Ber}(p)$ to denote a Bernoulli random variable with parameter $p \in [0, 1]$, and $X \sim \left[\text{Ber}(p)\right]^n$ to denote the distribution over $\mathbb{F}_2^n$ where each entry is i.i.d. Ber$(p)$.

1.6 Outline of article

Section 2 is a preliminary section that introduces symmetric sets in $\mathbb{F}_2^n$ and discusses tight asymptotic bounds on their sizes. Section 3 contains our results for P-LBC. Specifically, section 3.1 focuses on when the two classes are “separated”, and section 3.2 focuses on when they are not. Our P-SAS results are stated in section 4.1, along with various corollaries and remarks. These are proved in section 4.2. Section 5 contains directions for future work. The appendix contains several technical proofs.
2 Geometry of symmetric sets in $\mathbb{F}_2^n$

2.1 Preliminaries

The problems of LSC, SAS, and LBC are all stated for arbitrary sets in $\mathbb{F}_2^n$. However, much of the coding theory literature focuses specifically on symmetric sets because of their elegant analytic and geometric properties, their use in practical applications, and the fact that they model many sets that are efficiently representable [MacWilliams and Sloane 1977; van Lint 2012; Abbe et al. 2016]. Let us define symmetric sets formally.

**Definition 6** (Symmetric sets in $\mathbb{F}_2^n$). A set $S \subseteq \mathbb{F}_2^n$ is a symmetric set if for every $x \in S$, every permutation of $x$ is also in $S$.

A simple yet useful observation is that the set of symmetric sets is closed under Minkowski sums.

**Lemma 1.** If $S_1$ and $S_2$ are symmetric sets, then so is $S_1 + S_2$.

**Proof.** Take any element $x_1 + x_2$ of the sumset $S_1 + S_2$, and any permutation $\sigma \in \text{Sym}_n$. Because $S_1$ and $S_2$ are symmetric, $\sigma(x_1) \in S_1$ and $\sigma(x_2) \in S_2$. Thus $\sigma(x_1 + x_2) = \sigma(x_1) + \sigma(x_2) \in S_1 + S_2$. \qed

Common examples of symmetric sets we will work with in this article are Hamming balls and annuli, which are defined as follows.

**Definition 7** (Hamming ball, Hamming annulus). The Hamming ball centered around $x \in \mathbb{F}_2^n$ with radius $s \in \{0, \ldots, n\}$ is defined as:

$$B(x, s) = \{y \in \mathbb{F}_2^n : d(x, y) \leq s\}$$

For $0 \leq a \leq b \leq n$, the Hamming annulus $A(a, b, n)$ is defined as:

$$A(a, b, n) = \{y \in \mathbb{F}_2^n : w(y) \in [a, b]\}$$

In this article, we will study LSC, SAS, and LBC for general symmetric sets in $\mathbb{F}_2^n$, but most applications will be for Hamming balls and annuli. In particular, coding theory often focuses on Hamming balls and annuli centered around $0^n$, namely $B(0^n, s) = A(0, s, n)$. However, for certain problems such as SAS, we will want to remove the all-zero vector $0^n$, since no linear subspace can avoid that point. As such, it is often useful to consider punctured sets, which are sets that have the all-zero vector $0^n$ removed. In particular, the punctured Hamming ball centered around $0^n$ with radius $s$ is denoted by $B^*(0^n, s) = B(0^n, s) \setminus \{0^n\} = A(1, s, n)$.

With only these basic definitions, we can already give a few equivalent characterizations of symmetric sets that will prove useful. The intuition is that a symmetric set is completely characterized by a set of “permitted” Hamming weights, namely the Hamming weights of its elements.

**Remark 1.** For a set $S \subseteq \mathbb{F}_2^n$, the following are equivalent:

- $S$ is a symmetric set
- $S$ can be decomposed into the union of disjoint Hamming annuli
- There exists a set $T \subseteq \{0, \ldots, n\}$ such that $S = \{x \in \mathbb{F}_2^n : w(x) \in T\}$

2.2 Tight bounds on the size of symmetric sets

The proofs of our results in sections 3 and 4 will repeatedly rely on a few well-known facts about the sizes of Hamming balls and annuli. Before we state these, we first need to define a few basic information theoretic quantities. We refer the reader to MacKay 2003 and Cover and Thomas 2006 for a more detailed introduction.

**Definition 8** (Shannon entropy). Let $X$ be a random variable with probability mass function $P$. The Shannon entropy (over base $a$) of $X$ is defined to be $H_a(X) = \mathbb{E}[\log_a(P(x))] = -\sum_x P(x) \log_a P(x)$. 

7
Because this article focuses on finite fields of characteristic 2, we will primarily use the base 2 logarithm for entropy. So we adopt the convention that $H(X)$ denotes $H_2(X)$ when the base is unspecified.

A particularly useful quantity is the Shannon entropy of Bernoulli random variables. Observe that if $X \sim \text{Ber}(p)$ is a Bernoulli random variable with parameter $p \in (0,1)$, then its Shannon entropy is:

$$H(X) = -p \log p - (1 - p) \log(1 - p)$$  \hfill (8)

We often denote this quantity by $H(p)$ for shorthand. When we study the size of symmetric sets below, we will often refer to the adjusted Shannon entropy of $X \sim \text{Ber}(p)$, which is defined as:

$$\tilde{H}(X) = \begin{cases} 
H(p) & p \leq \frac{1}{2} \\
H\left(\frac{1}{2}\right) = 1 & p > \frac{1}{2}
\end{cases}$$  \hfill (9)

and is denoted by $\tilde{H}(p)$ for shorthand.

We are now ready to present well-known tight asymptotic bounds on the sizes of Hamming balls, Hamming annuli, and general symmetric sets.

**Lemma 2** (Size of Hamming ball). For any constant $\alpha \in [0,1]$:

$$|B(0^n, \alpha n)| = 2^{n\tilde{H}(\alpha)+o(n)}$$  \hfill (10)

**Proof.** $|B(0^n, \alpha n)| = \sum_{k=0}^{\alpha n} \binom{n}{k}$. Stirling’s approximation gives that $\ln(n!) = n \ln n - n + o(n)$, so for any $0 \leq k \leq n$:

$$\binom{n}{k} = 2^{n \log n - k \log k - (n-k) \log (n-k) + o(n)} = 2^{n\left(-\frac{k}{n} \log \frac{k}{n} - \frac{n-k}{n} \log \frac{n-k}{n}\right) + o(n)} = 2^{n\tilde{H}(\frac{k}{n}) + o(n)}$$  \hfill (11)

If $\alpha \in [0, \frac{1}{2}]$, then $2^{n\tilde{H}(\alpha)+o(n)} = \sum_{k=0}^{n\alpha} \binom{n}{k} \leq |B(0^n, \alpha n)| \leq (n\alpha + 1)\binom{n}{n\alpha} = 2^{n\tilde{H}(\alpha)+o(n)}$. Otherwise, $\alpha \in [\frac{1}{2}, 1]$, so $\tilde{H}(\alpha) = 1$. Thus $2^{n\tilde{H}(\alpha)+o(n)} = \sum_{k=0}^{n/2} \binom{n}{k} \leq |B(0^n, \alpha n)| \leq (n\alpha + 1)\binom{n}{n/2} = 2^{n\tilde{H}(\alpha)+o(n)}$, proving the claim. \hfill \Box

**Corollary 1** (Size of Hamming annulus). For any constants $0 \leq \alpha \leq \beta \leq 1$:

$$|A(\alpha n, \beta n, n)| = \begin{cases} 
2^{n\tilde{H}(\beta)+o(n)} & \alpha \leq \frac{1}{2} \\
2^{n\tilde{H}(1-\alpha)+o(n)} & \alpha > \frac{1}{2}
\end{cases}$$  \hfill (12)

**Proof.** By Lemma 2, $|A(\alpha n, \beta n, n)| = |B(0^n, \beta n)| - |B(0^n, \alpha n - 1)| = 2^{n\tilde{H}(\beta)+o(n)} - 2^{n\tilde{H}(\alpha)+o(n)}$. If $\alpha \leq \frac{1}{2}$ then this is dominated by $2^{n\tilde{H}(\beta)+o(n)}$. Otherwise $\frac{1}{2} \leq \alpha \leq \beta$, so $1 - \beta < \frac{1}{2}$. Thus by symmetry we can calculate $|A(\alpha n, \beta n, n)| = |A(1 - \beta n, (1 - \alpha)n, n)| = 2^{n\tilde{H}(1-\alpha)+o(n)} = 2^{n\tilde{H}(1-\alpha)+o(n)}$. \hfill \Box

In words, Corollary 1 says that the mass of an annulus in $F_2^n$ is concentrated around the vectors with Hamming weight equal closest to $\frac{n}{2}$. An identical result holds for general symmetric sets since we can decompose them into the union of disjoint annuli by Remark 1. This will become extremely useful for us in sections 3 and 4 when we analyze P-LBC and P-SAS for general symmetric sets. We formalize this idea algebraically in Corollaries 2 and 3 below, but to do so succinctly, we first define some notation to express the sizes of sets.

**Definition 9** (Relative weight, relative radius, relative size). Let $S \subseteq F_2^n$. Then the set of relative weights of $S$ is defined as:

$$W(S) = \left\{ \frac{w(v)}{n} : v \in S \right\}$$  \hfill (13)

The relative radius of $S$ is defined as:

$$\rho(S) = \max_{w \in W(S)} \frac{w}{n}$$  \hfill (14)
The relative size of $S$ is defined as:

$$\sigma(S) = \max_{w \in \mathcal{W}(S)} H(w)$$  \hfill (15)

If the limits exist, we define the relative radius and relative size of a family of sets $\mathcal{F} = \{S_n \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$ to be the limits $\rho(\mathcal{F}) = \lim_{n \to \infty} \rho(S_n)$ and $\sigma(\mathcal{F}) = \lim_{n \to \infty} \sigma(S_n)$, respectively.

**Example 1** (Relative size of a family of annuli and balls). Consider the family of Hamming annuli $\mathcal{F} = \{A(\alpha n, \beta n, n)\}_{n \in \mathbb{N}}$ where $0 \leq \alpha \leq \beta \leq 1$. The relative radius of any of these annuli is $\beta$ whereas the relative size of each of the annuli, and thus also of the family, is:

$$\sigma(A(\alpha n, \beta n, n)) = \sigma(\mathcal{F}) = \frac{1}{2} H(\beta)$$

In particular, letting $\alpha = 0$ makes $\mathcal{F}$ a family of Hamming balls with relative radius $\beta$ and relative size:

$$\sigma(B(0^n, \beta n)) = \sigma(\mathcal{F}) = \begin{cases} H(\beta) & \beta \leq \frac{1}{2} \\ H(\frac{1}{2}) = 1 & \beta \geq \frac{1}{2} \end{cases}$$

Thus $\sigma(\mathcal{F}) = \tilde{H}(\beta)$.

**Remark 2.** By symmetry of $H(\cdot)$ around $\frac{1}{2}$, the relative size $\sigma(S)$ can also be thought of as the entropy of the relative weight closest to $\frac{1}{2}$, i.e.:

$$\sigma(S) = H\left(\arg\min_{w \in \mathcal{W}(S)} \left| \frac{1}{2} - w \right| \right)$$

This is in contrast to the relative radius $\rho(S)$, which is simply the largest relative weight.

**Corollary 2** (Size of symmetric sets). If $S \subseteq \mathbb{F}_2^n$ if a symmetric set, then $|S| = 2^{-n\sigma(S) + o(n)}$.

**Proof.** By remark 2, $S$ can be decomposed as the union of disjoint annuli. The claim then follows by Corollary 1 and the definition of $\sigma(S)$. \hfill $\square$

**Corollary 3** (Size of a family of symmetric sets). If $\mathcal{F} = \{S_n \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$ is a family of symmetric sets with well-defined relative size, then $|S_n| = 2^{-n\sigma(\mathcal{F}) + o(n)}$.

**Proof.** Follows directly from Corollary 2 and Definition 9. \hfill $\square$
3 Results for Probabilistic Linear Boolean Classification

As discussed in subsection 1.2, one may only need to recover a “characteristic” or “classification” of a message instead of the exact message in full. Recall our motivating example for LBC: use the fewest linear queries to distinguish between mostly white images (elements of $S_{1,n} = B(0^n, \lceil an \rceil)$) and mostly black images (elements of $S_{2,n} = B(1^n, \lceil an \rceil)$). We mentioned that lossless LBC required a linear number $2\lceil an \rceil + 1$ of queries [Abbe et al. 2016]; but if we allow a vanishing probability of error in P-LBC, then the number of queries can grow to infinity arbitrarily slowly (e.g. $\log n$ or $\log \log \log \log \log \log n$). This new result is the the focus of section 3.1 below.

In fact, our result generalizes to any family of “separated” symmetric sets, which we will define formally shortly. This notion illustrates the intuition that: the farther apart that the closest elements of each set are, the easier it will be to classify a given vector since the MAP estimation will be less likely to err. As such, we separate our analysis of P-LBC on arbitrary families of symmetric sets into two regimes based on this measure of how “separated” these families are. Let us define these terms formally.

**Definition 10** (Inter-family distance). For all $n \in \mathbb{N}$, the $n^{th}$ inter-family distance between two families $F_1 = \{S_{1,n} \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$ and $F_2 = \{S_{2,n} \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$ is defined as:

$$d_n(F_1, F_2) = \min_{v_1 \in S_{1,n}, v_2 \in S_{2,n}} d(v_1, v_2)$$ (19)

We define “separated” and “tight” regimes based on whether the inter-family distance grows as $\Theta(n)$ or as $o(n)$, respectively. Formally:

**Definition 11** (Separated families, tight families). The families $F_1 = \{S_{1,n} \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$ and $F_2 = \{S_{2,n} \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$ are said to be:

- Separated families if $d_n(F_1, F_2) = \Theta(n)$.
- Tight families if $d_n(F_1, F_2) = o(n)$.

We split our study of P-LBC into classifying separated families and tight families. We analyze the former in section 3.1 and the latter in section 3.2.

We make one final note before proceeding to the technical sections below. In Definition 3, we phrased P-LBC with the message $X$ being drawn from a uniform mixture model, i.e. $X$ is drawn uniformly at random from $S_{1,n}$ with probability $\frac{1}{2}$, and otherwise drawn uniformly at random from $S_{2,n}$. However, we may also want to model when $X$ belongs to $S_{1,n}$ for some probability $p = \frac{1}{2}$. For example, returning to our example of mostly white images and mostly black images, it may be possible that mostly white images occur with probability 0.8 and mostly black images occur with probability 0.2. It is easy to generalize P-LBC to such types of non-uniform mixture models $X \sim_p (S_{1,n}, S_{2,n})$ for $p \in [0, 1]$, and we note that Lemma 3 and Theorem 1 below would still hold identically without any changes for this more general problem. We only stated the uniform mixture model for convenience and simplicity.

3.1 P-LBC for $\Theta(n)$ separated families: vanishing error probability with number of queries growing arbitrarily slowly

The main result of this section is the following.

**Theorem 1** (P-LBC for Separated Families). Let $F_1 = \{S_{1,n} \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$ and $F_2 = \{S_{2,n} \subseteq \mathbb{F}_2^n\}_{n \in \mathbb{N}}$ be any two separated families of symmetric sets. Then we can accomplish P-LBC with ranks $\{m_n\}_{n \in \mathbb{N}}$ that grow to infinity arbitrarily slowly.

In words, this means that with $m_n$ linear queries we may identify which of two sets a vector is in w.h.p., where $m_n$ is as small as we want provided $m_n \to \infty$ eventually as $n \to \infty$. For example, we could take the extremely slowly growing functions $m_n = \log \log \log \log \log \log n$ or even $m_n = \log^* \log^* \log^* \log^* n$.

Before presenting the proof of Theorem 1 we first state and prove it for the special case of Hamming annuli. The proof of Theorem 1 will then not be too difficult after we have proved this first, specialized result, given the connections between Hamming annuli and symmetric sets we have discussed in Remark 1.
Lemma 3 (P-LBC for Separated Annuli). For any $0 \leq a \leq b < c \leq d \leq n$, define the separated families of annuli $F_1 = \{ A(an, bn, n) \}_{n \in \mathbb{N}}$ and $F_2 = \{ A(cn, dn, n) \}_{n \in \mathbb{N}}$. Then we can accomplish P-LBC with ranks $\{m_n\}_{n \in \mathbb{N}}$ that grow to infinity arbitrarily slowly.

The proof strategy of Lemma 3 is quite simple: for each $n \in \mathbb{N}$, let $T_n$ be a random projection onto $m_n$ of the $n$ coordinates; and observe that the MAP estimation from the subsample $T_n(X)$ to decide which class $X$ is in precisely coincides with a majority vote. What we need to show is that this majority vote succeeds with high probability whenever the number of samples $m_n$ tends to $\infty$ as $n \to \infty$. We formalize this intuition as follows.

**Proof of Lemma 3** For all $n \in \mathbb{N}$, take $T_n : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^{m_n}$ to be a projection of $\mathbb{F}_2^n$ onto a random subset (with replacement, for ease of analysis below) of $m_n < n$ coordinates. The maximum-likelihood decoding scheme is then equivalent to declaring a message $X \in S_{1,n} \cup S_{2,n}$ to be in $S_{1,n}$ (resp. $S_{2,n}$) if less (resp. more) than $\frac{b+c}{2} m_n$ of the entries in $T_n(X)$ are 1, where $S_{1,n} = A(an, bn, n)$ are the elements of $F_1$ and $S_{2,n} = A(cn, dn, n)$ are the elements of $F_2$.

Recalling definition 3 it suffices to show that

$$\lim_{n \to \infty} \mathbb{P}_{X \sim \mu_n} \left\{ i(X) = \arg \max_{i \in \{1,2\}} \{ x' \in S_{i,n} : T_n(x') = T_n(X) \} \right\} = 1 \quad (20)$$

where $\mu_n$ denotes the uniform mixture model over the set $S_{1,n}$ and $S_{2,n}$ and $i(X) \in \{1,2\}$ is the random variable corresponding to whether $X$ belongs to the set $S_{1,n}$ or $S_{2,n}$.

By definition of $\mu_n$, $i(X)$ is either 1 or 2 with probability $\frac{1}{2}$. Thus for each $n \in \mathbb{N}$:

$$\mathbb{P}_{X \sim \mu_n} \left\{ i(X) = \arg \max_{i \in \{1,2\}} \{ x' \in S_{i,n} : T_n(x') = T_n(X) \} \right\} = \frac{1}{2} \sum_{j=1}^{2} \mathbb{P}_{X \sim S_{j,n}} \left\{ j = \arg \max_{i \in \{1,2\}} \{ x' \in S_{i,n} : T_n(x') = T_n(X) \} \right\} \quad (21)$$

Thus a necessary and sufficient condition for P-LBC is that for each $j \in \{1,2\}$:

$$\lim_{n \to \infty} \mathbb{P}_{X \sim S_{j,n}} \left\{ j = \arg \max_{i \in \{1,2\}} \{ x' \in S_{i,n} : T_n(x') = T_n(X) \} \right\} = 1 \quad (22)$$

We will show this for $j = 1$; an identical arguments holds for $j = 2$. We analyze the decoding-error probability using measure-concentration techniques. Recalling that $S_{1,n} = A(an, bn, n)$, we obtain:

$$\mathbb{P}_{X \sim A(an, bn, n)} \left\{ j = \arg \max_{i \in \{1,2\}} \{ x' \in S_{i,n} : T_n(x') = T_n(X) \} \right\} \quad (23)$$

$$= \mathbb{P}_{X \sim A(an, bn, n)} \left\{ w(T_n(X)) \leq \frac{b+c}{2} m_n \right\} \quad (24)$$

$$\geq \mathbb{P}_{X \sim A(cn, dn, n)} \left\{ w(T_n(X)) \leq \frac{b+c}{2} m_n \right\} \quad (25)$$

$$= \mathbb{P}_{Y \sim [\text{Ber}(b)]^n} \left\{ w(Y) \leq \frac{b+c}{2} m_n \right\} \quad (26)$$

$$\geq \left( 1 - e^{-2m_n \left( \frac{b+c}{2} \right)^2} \right) \quad (27)$$

Equation (26) is because the random subsamples are with replacement, so each entry of $T_n(X)$ has probability $\frac{b+c}{2}$ of being 1, i.e. is distributed $\text{Ber}(\frac{b+c}{2})$. Equation (27) is due to the well-known Hoeffding’s inequality [Boucheron et al., 2013]. Since $b < c$ are constants, $\lim_{n \to \infty} e^{-2m_n \left( \frac{b+c}{2} \right)^2} = 0$ for any sequence $\{m_n\}_{n \in \mathbb{N}}$ satisfying $\lim_{n \to \infty} m_n = \infty$. This completes the proof.

We are now ready to prove Theorem 1.
Proof. (Proof of Theorem 1) We have by Definition 11 that \( d_n(\mathcal{F}_1, \mathcal{F}_2) = \Theta(n) \). Thus there exists some constant \( \delta > 0 \) and sufficiently large \( N \) such that for every \( n \geq N \):
\[
\min_{v_1 \in S_{1,n}, v_2 \in S_{2,n}} d(v_1, v_2) = d_n(\mathcal{F}_1, \mathcal{F}_2) \geq \delta n \quad (28)
\]
For each such \( n \), we have by assumption that both \( S_{1,n} \) and \( S_{2,n} \) are symmetric sets; thus by Remark 1 they can each be decomposed into unions of disjoint annuli. By Corollary 1 the masses of \( S_{1,n} \) and \( S_{2,n} \) are concentrated in the annuli closest to Hamming weight \( \frac{1}{2} \). Formally, define \( S'_{1,n} = A(a_{\ell,n}, b_{\ell,n}, n) \cup A(a_{h,n}, b_{h,n}, n) \in S_{1,n} \) to be the union of the two maximal-by-inclusion annuli in \( S_{1,n} \) that have Hamming weight closest to \( \frac{1}{2} \), where \( a_{\ell,n} \leq b_{\ell,n} \leq \frac{1}{2} < a_{h,n} \leq b_{h,n} \). A similar construction for \( S_{2,n} \) finds \( S'_{2,n} = A(c_{\ell,n}, d_{\ell,n}, n) \cup A(c_{h,n}, d_{h,n}, n) \in S_{2,n} \), where \( c_{\ell,n} \leq d_{\ell,n} \leq \frac{1}{2} < c_{h,n} \leq d_{h,n} \). Importantly, Corollary 1 gives that \( \frac{|S'_{i,n}|}{|S_{i,n}|} = 1 - 2^{-\Theta(n)} \) for both \( i \in \{1, 2\} \), meaning that they are of the same size up to an exponentially small fraction.

Further, equation (25) gives that the absolute value of the difference between any element in \( \{a_{\ell,n}, b_{\ell,n}, a_{h,n}, b_{h,n}\} \) and any element in \( \{c_{\ell,n}, d_{\ell,n}, c_{h,n}, d_{h,n}\} \) is at least \( \delta \) for all sufficiently large \( n \). Therefore, using a constant number of size-\( m_n \) queries (using equation 27 in the proof of Lemma 3 above) between the two annuli components of \( S'_{1,n} \) and the two annuli components of \( S'_{2,n} \), we may with high probability determine which of the four annuli components \( X' \sim \frac{1}{2} (S'_{1,n}, S'_{2,n}) \) is in, and thus in particular whether \( X' \in S'_{1,n} \) or \( S'_{2,n} \). Since we have shown that \( \frac{|S'_{i,n}|}{|S_{i,n}|} = 1 - 2^{-\Theta(n)} \) for both \( i \in \{1, 2\} \), this implies that, with probability tending to 1 as \( n \to \infty \), we can determine whether \( X \sim \frac{1}{2} (S_{1,n}, S_{2,n}) \) is in \( S_{1,n} \) or \( S_{2,n} \) by a typical union bound.

3.2 P-LBC for \( o(n) \) tight families: simple upper bound and conjecture on lower bound

It is more difficult to do P-LBC when the families are tight instead of separated, as we had in the above section. This is understandable intuitively since by definition, tight families are “closer” to each other, which makes it harder to distinguish which class a given vector is in. We give partial results for this regime and leave as an open problem the asymptotically optimal P-LBC rank in Conjecture 1.

Instead of general symmetric sets (i.e. union of disjoint annuli), we instead focus simply on annuli in this section, since hopefully there should not be a large difference in the solution, given the results we found in Section 3.1 above. In fact, because of how easy it was to deal with separation in the previous section, we focus simply on annuli without width, i.e. recovering whether a given vector is in \( S_{1,n} = A(\alpha n, \alpha n, n) \) or \( S_{2,n} = A(\alpha n + c, \alpha n + c, n) \) for constants \( \alpha \in (0, 1) \) and \( c \in \mathbb{N}_+ \).

First note that a trivial parity-check suffices if \( c \) is odd.

Remark 3. If \( c \) is odd, then we can achieve zero error (i.e. we can accomplish WC-LBC and thus also P-LBC) with only 1 parity-check bit

Thus we are only concerned with \( c \) being even. But even the case \( c = 2 \) seems hard. In fact, we conjecture that it takes a linear number of queries to accomplish P-LBC:

Conjecture 1. P-LBC on \( \{A(\alpha n, \alpha n, n)\}_{n \in \mathbb{N}} \) and \( \{A(\alpha n + 2, \alpha n + 2, n)\} \) requires \( m_n = \Theta(n) \) for any \( \alpha \in (0, 1) \).

We now prove a simple upper bound by viewing P-LBC as instance of the harder problem of P-LC.

Theorem 2. Consider the families \( \{S_{1,n} = A(\alpha n, \alpha n, n)\}_{n \in \mathbb{N}} \) and \( \{S_{2,n} = A(\alpha n + c, \alpha n + c, n)\}_{n \in \mathbb{N}} \) for constants \( \alpha \in (0, 1) \) and \( c \in \mathbb{N}_+ \). Then P-LBC satisfies \( m_n \leq n H(\alpha) \) for all sufficiently large \( n \).

Proof. The proof idea is very simple. We rely on the well-studied results about “almost lossless linear source codes,” which are defined to be linear codes such that error of the decoding the entire message is vanishing. Clearly this is a harder task than P-LBC, since decoding the entire message will certainly guarantee decoding which class it is in. Therefore, the rate for P-LBC is upper bounded by the rate of this problem. The latter

\[2\text{When we refer to a “parity-check bit”, we mean the linear map } T : \mathbb{F}_2^n \to \mathbb{F}_2 \text{ defined by } T(x) = \sum_{i=1}^n x_i, \text{ where the addition is of course done over } \mathbb{F}_2.\]
is well-known from the probabilistic version of Shannon’s famous Source Coding Theorem for Symbol Codes [Shannon, 1948], which shows existence of almost lossless linear source codes \( \{T_n\}_{n \in \mathbb{N}} \) satisfying:

\[
\lim_{n \to \infty} \text{rank}(T_n) = \lim_{n \to \infty} H(D_n)
\]

where the \( \{D_n\}_{n \in \mathbb{N}} \) are the distributions over the source models. For us, \( D_n \) is the uniform mixture model over \( (S_{1,n}, S_{2,n}) \). Denote by \( D_{i,n} \) the uniform distribution over \( S_{i,n} \) for \( i \in \{1, 2\} \). By [Cover and Thomas, 2006], the entropy of a uniform distribution over a subset of \( \mathbb{F}_2^2 \) is the base-2 logarithm of its size; thus for both \( i \in \{1, 2\} \) we have

\[
H(D_{i,n}) = \log |S_{i,n}| = 2^{nH(\alpha) + o(n)}
\]

by Corollary 1. This allows us to calculate \( H(D_n) \) easily:

\[
H(D_n) = \sum_{x \in S_{1,n} \cup S_{2,n}} D_n(x) \log \frac{1}{D_n(x)} = \sum_{i=1}^{2} \sum_{x \in S_{i,n}} D_n(x) \log \frac{1}{D_n(x)} = \sum_{i=1}^{2} \sum_{x \in S_{i,n}} \frac{D_{i,n}(x)}{2} \log \frac{2}{D_{i,n}(x)}
\]

\[
= \frac{1}{2} \sum_{i=1}^{2} \left( \sum_{x \in S_{i,n}} D_{i,n}(x) \log \frac{1}{D_{i,n}(x)} + \sum_{x \in S_{i,n}} D_{n,i}(x) \log 2 \right) = \frac{1}{2} \sum_{i=1}^{2} (H(D_{i,n}) + 1)
\]

\[
= 1 + \frac{1}{2} \sum_{i=1}^{2} \log_2 2^{nH(\alpha) + o(n)} = O(nH(\alpha))
\]

In words, the above proof uses the reduction that probabilistic LSC is at least as hard as P-LBC. We note that a result on Conjecture 1 stated above, would illuminate how much easier P-LBC is than probabilistic LSC, for these hard inputs of tight families. (Our result in Theorem 1 shows that P-LBC is in fact much easier than probabilistic LSC for separated families.)
4 Results for Probabilistic Subspace-Avoiding-Set Problem

In section 4.1, we state a lower bound for the $\gamma$-P-SAS capacities of general symmetric sets in $\mathbb{F}_2^n$, and then present a few corollaries and remarks as direct applications. We present the proof of the lower bound in section 4.2.

4.1 Lower bound on $\gamma$-P-SAS capacities for symmetric sets: statement, corollaries, and intuition

The main result of this section is a lower bound on the $\gamma$-P-SAS capacities of general symmetric sets in terms of their relative sizes. In words, this shows the existence of large subspaces that “mostly” avoid given symmetric sets. Our lower bound for the dimension of these subspaces is given as a function of the allowed number of intersections (errors), which is in turn controlled by the error parameter $\gamma \in [0, 1]$. This is formally stated as follows.

**Theorem 3** (P-SAS-capacity for symmetric sets). Let $\mathcal{F} = \{ S_n \in \mathbb{F}_2^n \setminus \{0^n\} \}_{n \in \mathbb{N}}$ be a family of symmetric sets with well-defined relative size $\sigma(\mathcal{F})$. For any $\gamma \in [0, 1]$, the supremum of $\gamma$-achievable rates is at least:

$$k^*_\gamma(\mathcal{F}) \geq 1 - (1 - \gamma)\sigma(\mathcal{F})$$

(33)

This immediately gives a lower bound on $\gamma$-P-SAS capacities for Hamming balls and annuli since they are special cases of symmetric sets. Recalling the characterization of $\sigma(\mathcal{F})$ for such sets in Example 1 gives the following two corollaries.

**Corollary 4** (P-SAS-capacity for Hamming annuli). For any constants $0 \leq \alpha \leq \beta \leq n$, let $\mathcal{F} = \{ A(\alpha n, \beta n, n) \}_{n \in \mathbb{N}}$. Then for any $\gamma \in [0, 1]$:

$$k^*_\gamma(\mathcal{F}) \geq 1 - (1 - \gamma)\left\{ \begin{array}{ll} H(\beta) & \alpha, \beta \leq \frac{1}{2} \\ 1 & \alpha \leq \frac{1}{2} \leq \beta \\ H(\alpha) & \frac{1}{2} \leq \alpha, \beta \end{array} \right.$$  

(34)

**Corollary 5** (P-SAS-capacity for Hamming balls). For any constant $\alpha \in [0, 1]$, let $\mathcal{F} = \{ B^*(0^n, [\alpha n]) \}_{n \in \mathbb{N}}$ denote the family of punctured Hamming balls with relative radius $\alpha$. Then for any $\gamma \in [0, 1]$:

$$k^*_\gamma(\mathcal{F}) \geq 1 - (1 - \gamma)\bar{H}(\alpha)$$

(35)

Recall from section 1.3 that P-SAS is defined with the parameter $\gamma \in [0, 1]$ to allow us to understand how the large the subspaces we can find are as a function of the error we are allowed to make. As we mentioned, both the settings $\gamma = 0$ and $\gamma = 1$ have natural interpretations: the former is WC-SAS (which has Dual WC-LSC as special case); and the latter requires that a randomly chosen vector in the set will lie in our subspace with vanishing probability. We discuss what Theorem 3 says about these settings $\gamma \in \{0, 1\}$ in the following two remarks, and note that by equation (33), the $\gamma$-P-SAS capacity interpolates linearly between these two extremes when the parameter $\gamma$ varies within $[0, 1]$.

**Remark 4.** When $\gamma = 0$, the lower bound in Corollary 3 gives a lower bound of $n(1 - \bar{H}(\alpha))$ for Dual WC-LSC, and thus an upper bound of $n\bar{H}(\frac{\alpha}{2})$ for Primal WC-LSC. This matches the well-known Gilbert-Varshamov bound [Varshamov, 1957]. This means that our bound is tight for $\gamma = 0$ if Goppa’s conjecture is true [Goppa, 1993].

**Remark 5.** Plugging $\gamma = 1$ into Theorem 3 shows that we can find subspaces $X_n$ of arbitrarily large dimension $n + o(n)$ such that $P_{Y \sim S_n}(Y \in X_n) \to 0$ as $n \to \infty$.

To give intuition for the seemingly paradoxical statement in Remark 5, observe that if $S_n = \mathbb{F}_2^n$, then subspaces $X_n$ of dimension $(1 - \varepsilon)n$ satisfy $P_{Y \sim S_n}(Y \in X_n) = \frac{|Y \cap S_n|}{|S_n|} = 2^{-n\varepsilon} \to 2^{-\varepsilon n} \to 0$. We also give a simple calculation that gives intuition for when $S_n$ is not so degenerate: consider $S_n = B^*(0^n, [\alpha n])$ the punctured Hamming ball with relative radius $\alpha \in [0, 1]$. To see that the 1-P-SAS capacity for these

14
balls is 1, we show that every $c < 1$ is a 1-P-SAS achievable rate. To do this, define the linear subspace $\{X_n = \text{span}\{e_1, \ldots, e_{\lceil cn \rceil}\}\}$. Then by Lemma 2,

$$\frac{|X_n \cap B^*(0^n,\lceil cn \rceil)|}{|B^*(0^n,\lceil cn \rceil)|^\gamma} = \frac{|B^*(0,\lceil cn \rceil)|}{|B^*(0^n,\lceil cn \rceil)|^\gamma} = \frac{2^{|\lceil cn \rceil| H(\tilde{\alpha})}}{2^{\gamma n H(\alpha)}} \cdot 2^{\sigma(n)}$$

(36)

This tends to 0 if and only if:

$$|cn| H\left(\frac{\alpha}{c}\right) - \gamma n H(\alpha) \rightarrow -\infty \iff c H\left(\frac{\alpha}{c}\right) - \gamma H(\alpha) < 0 \iff \tilde{H}\left(\frac{\alpha}{c}\right) < \gamma H(\alpha)$$

(37)

Note that when $\gamma = 1$, this holds for any $c < 1$ by Jensen’s inequality, since $\tilde{H}(\cdot)$ is concave. Thus every $c < 1$ is a 1-P-SAS achievable rate by definition, and so 1 is the 1-P-SAS capacity, as desired.

### 4.2 Proof of lower bound via probabilistic method

**Proof outline.** To prove Theorem 3, it suffices to show that any $c < 1 - (1 - \sigma(\mathcal{F}))$ is a $\gamma$-P-SAS achievable rate for $\mathcal{F}$. We prove this with the probabilistic method. Specifically, we will take a distribution over $\lceil cn \rceil$-dimensional subspaces, and show that with strictly positive probability, a randomly drawn subspace (according to this distribution) satisfies the desired the property for $\gamma$-P-SAS. This then implies the existence of a desired $\lceil cn \rceil$-dimensional subspace.

Let us formalize this argument slightly. We need to find a sequence of $\lceil cn \rceil$-dimensional subspaces satisfying

$$\lim_{n \rightarrow \infty} \frac{|S_n \cap X_n|}{|S_n|^\gamma} = 0$$

(38)

A key trick will be to rewrite:

$$\frac{|S_n \cap X_n|}{|S_n|^\gamma} = \frac{|X_n|}{|S_n|^\gamma} \left(\frac{|X_n \cap S_n|}{|X_n|}\right)$$

(39)

The left fraction can be calculated using the simple fact that $|X_n| = 2^{\dim(X_n)}$ since $X_n$ is a subspace, and Corollary 2 to approximate the size of the symmetric set $|S_n|$. The right fraction is slightly more difficult to analyze, but has a nice geometric interpretation: it is the probability that a randomly chosen vector in $X_n$ lies in $S_n$, i.e.:

$$\mathbb{P}_{Y \sim X_n}\{Y \in S_n\} = \frac{|X_n \cap S_n|}{|X_n|}$$

(40)

Informally, we must show that this probability is “small” in order for the term in equation (38) to tend to 0. This is where we use the probabilistic method to find such a sequence of “good” subspaces $X_n$. Specifically, we draw vectors $Y_1, \ldots, Y_{\lceil cn \rceil}$ uniformly at random from $\mathbb{F}_2^n$ and define the random vector space $X_n = \text{span}\{Y_1, \ldots, Y_{\lceil cn \rceil}\}$. Our first step is to show that with high probability, $Y_1, \ldots, Y_{\lceil cn \rceil}$ are linearly independent (and thus $X_n$ is in fact of dimension $\lceil cn \rceil$). Next we show that, conditional on this high probability event, the expectation $\mathbb{E}_{X_n}[\mathbb{P}_{Y \sim X_n}\{Y \in S_n\}]$ is “small” using measure-concentration techniques. By the probabilistic method, this guarantees the existence of a subspace $X_n$ that makes the probability term in equation (38) “small”, which allows us to conclude the proof by plugging in to equation (39) and then equation (38).

**Proofs of lemmas** We now embark on proving the lemmas and partial results mentioned above that we will need to prove Theorem 3.

The first lemma lower bounds the probability that randomly drawn vectors $Y_1, \ldots, Y_k$ are linearly independent. We note that this appears as Lemma 1 of Altschuler and Yang [2015], but we give its proof for completeness since it is short.
Lemma 4 (Lemma 1 of [Altschuler and Yang 2015]).

\[ P_{Y_1, \ldots, Y_k \sim \mathbb{F}_2^n} \{ Y_1, \ldots, Y_k \text{ are linearly independent} \} > 1 - 2^{k-n} \]  

(41)

Proof. By the chain rule and a typical union bound:

\[ P_{Y_1, \ldots, Y_k \sim \mathbb{F}_2^n} \{ Y_1, \ldots, Y_k \text{ are LI} \} = \prod_{i=1}^{k} P_{Y_i \sim \mathbb{F}_2} \{ Y_i \notin \text{span}\{Y_1, \ldots, Y_{i-1}\} \ | \ Y_1, \ldots, Y_{i-1} \text{ are LI} \} \]

(42)

\[ = \prod_{i=1}^{k} 1 - 2^{i-1-n} \]  

(43)

\[ \geq 1 - \sum_{i=1}^{k} 2^{i-1-n} \]  

(44)

\[ > 1 - 2^{k-n} \]  

(45)

\[ \square \]

The following two lemmas show the equivalence of certain distributions, which we will make use of below. Because their proofs are slightly technical and not particularly insightful, we defer them to the appendix (see sections A.1 and A.2).

Lemma 5. Let \( Y_0, Y_1, \ldots, Y_k \sim \text{Ber}(\frac{1}{2}) \) for \( k \geq 0 \). Then the following two distributions are equivalent:

1. \( Y = Y_0 + (C_1 Y_1 + \ldots C_k Y_k) \) for \( \{ C_i \}_{i=1}^k \sim \text{Ber}(\frac{1}{2}) \)

2. \( Y \sim [\text{Ber}(\frac{1}{2})]^n \)

Lemma 6. Fix \( y_1, \ldots, y_k \in \mathbb{F}_2^n \). Then the following two distributions are equivalent:

1. \( Y \sim \text{span}\{y_1, \ldots, y_k\} \)

2. \( Y = C_1 y_1 + \ldots C_k y_k \) for \( C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2}) \)

The next step is to show that \( \mathbb{E}_n[ P_{Y \sim X_n} \{ Y \in S_n \} ] \) is “small”. As we mentioned earlier, this will guarantee the existence of a subspace \( X_n \) such that \( P_{Y \sim X_n} \{ Y \in S_n \} \) is “small”, which will be an essential step in the overall proof.

Lemma 7. Let \( Y_1, \ldots, Y_k \sim [\text{Ber}(\frac{1}{2})]^n \), and define \( X = \text{span}\{Y_1, \ldots, Y_k\} \). Then for all \( n \in \mathbb{N} \):

\[ \mathbb{E}_X[ P_{Y \sim X} \{ Y \in S_n \} ] = 2^{-n(1-\sigma(S_n))+o(n)} \cdot \left( 1 - \frac{1}{2k} \right) \]  

(46)

Proof. Calculations show:

\[ \mathbb{E}_X[ P_{Y \sim X} \{ Y \in S_n \} ] \]

(47)

\[ = \mathbb{E}_Y, Y \sim [\text{Ber}(\frac{1}{2})]^n [ P_{Y \sim \text{span}\{Y_1, \ldots, Y_k\}} \{ Y \in S_n \} ] \]

(48)

\[ = \mathbb{E}_Y, Y \sim [\text{Ber}(\frac{1}{2})]^n [ P_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})} \{ \sum_{i=1}^{k} C_i Y_i \in S_n \} ] \]

(49)

\[ = \mathbb{E}_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})} \mathbb{E}_Y, Y \sim [\text{Ber}(\frac{1}{2})]^n [ \{ \sum_{i=1}^{k} C_i Y_i \in S_n \} ] \]

(50)

\[ = \mathbb{E}_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})} \mathbb{E}_Y, Y \sim [\text{Ber}(\frac{1}{2})]^n [ \{ \sum_{i=1}^{k} C_i Y_i \in S_n \} \ | \ \exists C_i \neq 0 \} \cdot P_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})} \{ \exists C_i \neq 0 \} \]

(51)

\[ = \mathbb{E}_{Y \sim [\text{Ber}(\frac{1}{2})]^n} [ \{ Y \in S_n \} ] \cdot \left( 1 - \frac{1}{2k} \right) \]

(52)
Equation (49) is due to Lemma 6. Equation (50) is due to Fubini’s Theorem. Equation (51) is from conditioning on the event that $C_i = 0$ for all $i$; note that in this event, obviously $C_1 Y_1 + \cdots + C_k Y_k = 0^n \notin S_n$ since $S_n \subseteq \mathbb{F}_2 \setminus \{0^n\}$ by assumption. Equation (52) is due to Lemmas 3 and 6.

Now we need to analyze the large-deviation term $\mathbb{E}_{Y \sim \text{Ber}(\frac{1}{2})} \left[ 1 \left( Y \in S_n \right) \right]$. Sanov’s Theorem will give a tight concentration bound [Sanov 1958]. Denote the coordinates of $Y$ by $Y = (Z_1, \ldots, Z_n)^T$, then:

$$
\mathbb{P}_{Y \sim \text{Ber}(\frac{1}{2})} \left\{ Y \in S_n \right\} = \mathbb{P}_{Z_1, \ldots, Z_n \sim \text{Ber}(\frac{1}{2})} \left\{ \sum_{i=1}^n Z_i \in \mathcal{W}(S_n) \right\}
$$

$$
= e^{-n\inf_{x \in \mathcal{W}(S_n)} \left\{ D_{KL}(\text{Ber}(x) \mid \text{Ber}(\frac{1}{2})) \right\} + o(n)}
$$

$$
= e^{-n\inf_{x \in \mathcal{W}(S_n)} \left\{ \ln(2) - H_x(\alpha) \right\} + o(n)}
$$

$$
= e^{-n(1 - \alpha(S_n)) + o(n)}
$$

$$
= 2^{-n(1 - \sigma(S_n)) + o(n)}
$$

where we use $\sigma(\cdot)$ to denote $\sigma(\cdot)$ but with logarithm base $e$ instead of logarithm base 2 for the Shannon entropy. Equation (54) is a direct application of Sanov’s Theorem, and equation (55) is due to Definition 3. Equation (55) is due to the basic identity from information theory that the KL-divergence:

$$
D_{KL}(P(X) \mid P_{\text{uniform}}(X)) = \ln(|X|) - H_x(X)
$$

The proof statement now follows directly from combining Equations (52) and (57).

We are now ready to conclude the existence of a “good” subspace $A$, which we formalize as follows using the probabilistic method.

**Lemma 8.** For any $n \in \mathbb{N}$ and $k \leq n$, there exists a $k$-dimensional linear subspace $A \subseteq \mathbb{F}_2^n$ satisfying:

$$
\mathbb{P}_{Y \sim A} \left\{ Y \in S_n \right\} < \frac{2^{-n(1 - \sigma(S_n)) + o(n)} \cdot \left( 1 - \frac{1}{2^n} \right)}{1 - \frac{2^k}{2^n}}
$$

**Proof.** By Lemma 7 and conditioning on the dimension of $X$,

$$
2^{-n(1 - \sigma(S_n)) + o(n)} \cdot \left( 1 - \frac{1}{2^n} \right) = \mathbb{E}_{X} \left[ \mathbb{P}_{Y \sim X} \left\{ Y \in S_n \right\} \right]
$$

$$
> \mathbb{E}_{X} \left[ \mathbb{P}_{Y \sim X} \left\{ Y \in S_n \mid \dim(X) = k \right\} \cdot \mathbb{P}\left( \dim(X) = k \right) \right]
$$

$$
> \mathbb{E}_{X} \left[ \mathbb{P}_{Y \sim X} \left\{ Y \in S_n \mid \dim(X) = k \right\} \cdot \left( 1 - \frac{2^k}{2^n} \right) \right]
$$

Equation (61) is due to conditioning on the event that $\dim(X) = k$. Equation (62) is due to Lemma 4 and Fubini’s theorem. Finally, note that equation (62) directly gives the existence of some $k$-dimensional subspace $A$ satisfying equation (59), completing the proof.

We are now finally ready to prove Theorem 3.

**Proof of Theorem 3**. Fix $c \in [0, 1)$. By Lemma 8, there exist a sequence of $\lceil cn \rceil$ dimensional subspaces $X_n \subseteq \mathbb{F}_2^n$ satisfying equation (59). Then:

$$
\frac{|X_n \cap S_n|}{|S_n|^\gamma} = \frac{|X_n|}{|S_n|^\gamma} \cdot \frac{|X_n \cap S_n|}{|X_n|}
$$

$$
= \frac{2^{\lceil cn \rceil}}{2^{\gamma n \sigma(S_n) + o(n)}} \mathbb{P}_{Y \sim X_n} \left\{ y \in S_n \right\}
$$

$$
< \frac{2^{\lceil cn \rceil}}{2^{\gamma n \sigma(S_n) + o(n)}} \cdot \frac{2^{-n(1 - \sigma(S_n)) + o(n)} \cdot \left( 1 - \frac{1}{2^n} \right)}{1 - \frac{2^{\lceil cn \rceil}}{2^n}}
$$

$$
= 2^{n \sigma(S_n)(1 - \gamma) + o(n)} \cdot \frac{2^{\lceil cn \rceil} - 1}{2^n - 2^{\lceil cn \rceil}}
$$

17
Thus $c \in [0,1)$ is a $\gamma$-P-SAS achievable rate iff (66) tends to 0 as $n$ tends to $\infty$. This happens iff:

$$\lim_{n \to \infty} 2^n \sigma(S_n)(1-\gamma) + o(n) \cdot \frac{2^c}{2^n} = 0$$

(67)

$$\lim_{n \to \infty} n \sigma(S_n)(1-\gamma) + cn - n = -\infty$$

(68)

$$\sigma(F)(1-\gamma) + c - 1 < 0$$

(69)

where equation (69) is due to the fact that $F = \{S_n\}_{n \in \mathbb{N}}$ is assumed to have well-defined relative size, i.e. $\sigma(F) = \lim_{n \to \infty} \sigma(S_n)$ is well-defined. Thus, by definition (5) every $c < 1 - (1-\gamma)\sigma(F)$ is a $\gamma$-P-SAS achievable rate for $\gamma \in [0,1]$. Therefore we can lower bound the $\gamma$-P-SAS capacity of $F$ by:

$$k_\gamma^*(F) = \sup_{c \in [0,1]} \{ c : c \text{ is a } \gamma\text{-P-SAS achievable rate for } F \} \geq 1 - (1-\gamma)\sigma(F)$$

(70)

completing the proof.
5 Future work

We have solved P-LBC for $\Theta(n)$ separated families in Theorem 1 but the problem is still open for $o(n)$ tight families. Our proposed Conjecture 1 is that any P-LBC decoding scheme for non-degenerate tight families with vanishing error probability must have number of queries scaling linearly with $n$. We have given a simple upper bound for this regime in Theorem 2 but a tighter upper bound as well as a lower bound would shed light on whether P-LBC is actually an easier problem than Shannon’s lossy source coding for these hard inputs – and if so, by how much.

An interesting extension of LBC is “sequence coding;” consider LBC between $k$ sets instead of 2. That is, given a source model $S = S_1 \cup \cdots \cup S_k$, encode a vector $x \in S$ such that it is possible to recover the $S_i$ in which $x$ lies. This could be analyzed in both worst-case (lossless) and probabilistic (lossy) settings. By our result in Theorem 1 we may determine whether $x$ is in an $S_i$ nearly “for free” if the set $S_i$ is separated from $S \setminus S_i$ by $\Theta(n)$ Hamming distance. A further extension would entail exact recovery of $x$ if $x$ was in some $S_i$, and only recovery of the fact that $x \in S_i$ if it was in one of the others. This second extension might find practical applications in settings where compression precision is dependent on how important the set to compress is.

We have given a lower bound for the $\gamma$-P-SAS capacity of general symmetric sets in Theorem 3 but its tightness is not clear. It is likely that our lower bound is tight at least for $\gamma = 0$ since that regime agrees with the Gilbert-Varshamov bound, which is conjectured to be tight (see Goppa’s conjecture [Goppa, 1993]). We believe that it is possible to obtain an information theoretic upper bound using new variants of Fano’s inequality. This would shed further light on how the difficulty of P-SAS changes as $\gamma$ varies.
6 Bibliography


A Appendix

A.1 Proof of Lemma 5

Proof. Every entry of $Y_0 + \sum_{i=1}^k C_i Y_i$ is i.i.d., since we can simply condition on the $C_i$ and remember that each $Y_i$ has i.i.d. entries. Note the importance of having a $Y_0$ without coefficients for this because otherwise the case $C_1 = \cdots = C_k = 0$ would be problematic. Next, observe that each entry is distributed $Ber(\frac{1}{2})$ since for any $j$: $(Y_0 + \sum_{i=1}^k C_i Y_{i,j}) = Y_{0,j} + \sum_{i=1}^k C_i Y_{i,j}$ is the sum of i.i.d. $Ber(\frac{1}{2})$ random variables (i.e. $Y_0$ and the $c_i Y_{i,j}$ for $C_i = c_i$ nonzero after conditioning), and so is also distributed $Ber(\frac{1}{2})$.

A.2 Proof of Lemma 6

Proof. We prove by induction on $k$. The base case is $k = 1$; two cases follow. If $y_1 = 0^n$, then $\text{span}\{y_1\} = \{0^n\}$, and so both distributions choose $0^n$ with probability 1. Otherwise $y_1 \neq 0^n$, so $\text{span}\{y_1\} = \{0^n, y_1\}$, and thus both distributions choose $0^n$ with probability $\frac{1}{2}$ and otherwise $y_1$.

Now the inductive step: assume true for $1, \ldots, k$, and we show true for $k+1$. We do casework on whether $y_{k+1} \in \text{span}\{y_1, \ldots, y_k\}$.

- Case 1: Assume $y_{k+1} \notin \text{span}\{y_1, \ldots, y_k\}$. We know $|\text{span}\{y_1, \ldots, y_k\}| = 2^m$ for some $m \leq k$, and thus $|\text{span}\{y_1, \ldots, y_{k+1}\}| = 2^{m+1}$. Fix any $y \in \text{span}\{y_1, \ldots, y_{k+1}\}$. Then the projections $y_{\text{span}\{y_1, \ldots, y_k\}}$ and $y_{\text{span}\{y_{k+1}\}}$ of $y$ onto the subspaces $\text{span}\{y_1, \ldots, y_k\}$ and $\text{span}\{y_{k+1}\}$, respectively, are well defined, so:

$$\mathbb{P}_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})}\{C_1 y_1 + \cdots + C_k y_k + y_{k+1} = y\} = \frac{1}{2m+1}$$

where equation (73) was by independence of the $C_i$ and equation (74) was by the strong induction hypothesis.

- Case 2: Assume $y_{k+1} \in \text{span}\{y_1, \ldots, y_k\}$. Then there exist coefficients $a_1, \ldots, a_k$ satisfying $\sum_{i=1}^k a_i y_i = y_{k+1}$. Let $S = \text{span}\{y_1, \ldots, y_{k+1}\} = \text{span}\{y_1, \ldots, y_k\}$. Then we have $|S| = 2^m$ for some $m \leq k$. Fix any $y \in S$. By the inductive hypothesis,

$$\mathbb{P}_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})}\{C_1 y_1 + \cdots + C_k y_k = y\} = \mathbb{P}_{Y \sim S}\{Y = y\} = \frac{1}{2^m} = \frac{2^{k-m}}{2^k}$$

Thus, there are $2^{k-m}$ distinct sets of coefficients $\{c_1, \ldots, c_k\}$ s.t. $c_1 x_1 + \cdots + c_k x_k = z$. Therefore:

$$\mathbb{P}_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})}\{C_1 y_1 + \cdots + C_k y_k + y_{k+1} = y\} = \frac{1}{2^m} \cdot \left[ \mathbb{P}_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})}\{C_1 y_1 + \cdots + C_k y_k = y\} + \mathbb{P}_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})}\{C_1 y_1 + \cdots + C_k y_k + y_{k+1} = y\} \right]$$

$$= \frac{1}{2^m} \cdot \left[ \mathbb{P}_{Y \sim \text{span}\{y_1, \ldots, y_k\}}\{Y = y\} + \mathbb{P}_{C_1, \ldots, C_k \sim \text{Ber}(\frac{1}{2})}\{C_1 y_1 + \cdots + (C_k + a_k) y_k = y\}\right]$$

$$\mathbb{P}_{Y \sim S \sim \text{span}\{y_1, \ldots, y_{k+1}\}}\{Y = y\}$$